



1-1-2014

Three Essays on Crime and Delinquency: Immigration, Alcohol, and Probation

SeungHoon Han

University of Pennsylvania, seunghh77@hotmail.com

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

Recommended Citation

Han, SeungHoon, "Three Essays on Crime and Delinquency: Immigration, Alcohol, and Probation" (2014). *Publicly Accessible Penn Dissertations*. 1307.

<http://repository.upenn.edu/edissertations/1307>

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/edissertations/1307>

For more information, please contact libraryrepository@pobox.upenn.edu.

Three Essays on Crime and Delinquency: Immigration, Alcohol, and Probation

Abstract

This dissertation is a collection of three distinct chapters. The first chapter explores the associations between various assimilation measures and outcomes of delinquent risk factor development for immigrant youth from Los Angeles County, CA, using both logit regression and doubly robust estimation methods. Youth with a low level of assimilation were more likely to have a high level of attachment to family and school than those with a high level of assimilation. However, the associations were less evident for the other delinquent risk factors. Among the assimilation measures used, the generation status measure was not predictive of the outcomes. The two language-related assimilation measures were less strongly associated with the outcomes than the immigrant status measure.

The second chapter investigates crime impacts of the increased alcohol availability made through repealing a Sunday off-premise liquor sales ban. Since 2003, Pennsylvania permitted a part of its state-run liquor stores to open on Sunday by repealing the ban, which enables a quasi-experimental triple difference design. Some evidence of local crime pattern changes after the repeal was found. The repeal was associated with an increase in total crime incidents occurring in the immediate vicinity of the Sunday-open liquor stores in Philadelphia. At the same time, total crime incidents occurring relatively farther away from the stores decreased. These pattern changes were present in low socioeconomic status (SES) neighborhoods, but not evident in high SES ones.

The third chapter evaluates whether the effect of a cognitive behavioral therapy (CBT) program on recidivism differ depending on probationer characteristics from a Bayesian perspective. Using Philadelphia CBT randomized controlled experiment data with a Bayesian hierarchical Gamma-Poisson model, the study compares average recidivism rates between the CBT and non-CBT groups, conditional on probationer characteristics. The Bayesian analysis showed that the effects of the CBT program were statistically meaningfully more evident for the high-risk probationers who were between 10-19 and 30-39 years old, who had more extensive prior experience on probation, and who had a higher ratio of high risk predictions. These results contradict the frequentist evaluation results that the CBT program had no statistically significant effect on recidivism.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Criminology

First Advisor

John M. MacDonald

Keywords

alcohol, assimilation, Bayesian, displacement, immigration, probation

Subject Categories

Criminology | Criminology and Criminal Justice

THREE ESSAYS ON CRIME AND DELINQUENCY:
IMMIGRATION, ALCOHOL, AND PROBATION

SeungHoon Han

A DISSERTATION

in

Criminology

Presented to the Faculties of the University of Pennsylvania

In

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

Graduate Group Chairperson

Dr. John M. MacDonald
Associate Professor of
Criminology and Sociology

Dr. Charles Loeffler
Jerry Lee Assistant Professor
of Criminology

Dissertation Committee:

Charles C. Branas	Professor of Epidemiology and Biostatistics
Philip J. Cook	ITT/Terry Sanford Professor of Public Policy, Professor of Economics and Sociology, Duke University
Richard J. Gelles	The Joanne and Raymond Welsh Chair of Child Welfare and Family Violence, Professor of Social Policy & Practice
Emily G. Owens	Associate Professor of Criminology

THREE ESSAYS ON CRIME AND DELINQUENCY
: IMMIGRATION, ALCOHOL, AND PROBATION

COPYRIGHT ©

2014

SeungHoon Han

DEDICATION

Dedicated to all my family and friends....

ACKNOWLEDGMENT

I have many people to whom I am indebted for this dissertation. Without them, I could never have completed it. John MacDonald, my advisor. I would like to express my special appreciation to you. Always encouraging me to be devoted to research and offering appropriate advice and guidance, you helped me to grow as a researcher. You have also been a good friend in life, not only my role model as a scholar. Philip Cook. You have always been my sincere mentor since I was a master student (and even before I came to the US, through your book!). I cannot thank you enough for the invaluable support you gave me for that long time. I also thank you for introducing to me the world of criminology, and for recognizing my potential as a researcher. Also, my gratitude to the committee members, Charles Branas, Richard Gelles, and Emily Owens. Your insightful comments and questions, along with the encouragement and support you showed, were extremely helpful in completing this dissertation.

To my dear program colleagues. Jordan Hyatt. I thank you for being my good friend and academic colleague (and (in)voluntarily an English tutor of mine, too!). You willingly shared many priceless experiences, for which I am deeply grateful. Also, Jill Portnoy, Frances Chen, Ellen Donnelly, and Ben Grunwald. I really enjoyed spending my PhD student life with you, including not only academic stimuli from discussion but also small chats for relaxation. In addition, Knaykia Hagans, and Janel Baslice. I sincerely appreciate the warmth you displayed. It was far more than just administrative support.

To my Korean friends. There are far too many names to list here. If you are a friend who happens to read this statement, I would like to say that I am grateful to you.

I would also like to thank the Business Improvement District (BID) team in the Rand Corporation, the Philadelphia Police Department, the Pennsylvania Liquor Control Board, the Philadelphia Adult Probation and Parole Department (APPD), and the Jerry Lee Center of Criminology (JLC) at the University of Pennsylvania for the data used herein.

Finally, to my family, and to my beloved wife, HyeSun. I cannot thank you enough for your endless love, financial and emotional support, and voluntary sacrifices. Without you, I would not be able to finish this long journey. I love you.

ABSTRACT

THREE ESSAYS ON CRIME AND DELINQUENCY : IMMIGRATION, ALCOHOL, AND PROBATION

SeungHoon Han

John M. MacDonald

This dissertation is a collection of three distinct chapters. The first chapter explores the associations between various assimilation measures and outcomes of delinquent risk factor development for immigrant youth from Los Angeles County, CA, using both logit regression and doubly robust estimation methods. Youth with a low level of assimilation were more likely to have a high level of attachment to family and school than those with a high level of assimilation. However, the associations were less evident for the other delinquent risk factors. Among the assimilation measures used, the generation status measure was not predictive of the outcomes. The two language-related assimilation measures were less strongly associated with the outcomes than the immigrant status measure.

The second chapter investigates crime impacts of the increased alcohol availability made through repealing a Sunday off-premise liquor sales ban. Since 2003, Pennsylvania permitted a part of its state-run liquor stores to open on Sunday by repealing the ban, which enables a quasi-experimental triple difference design. Some

evidence of local crime pattern changes after the repeal was found. The repeal was associated with an increase in total crime incidents occurring in the immediate vicinity of the Sunday-open liquor stores in Philadelphia. At the same time, total crime incidents occurring relatively farther away from the stores decreased. These pattern changes were present in low socioeconomic status (SES) neighborhoods, but not evident in high SES ones.

The third chapter evaluates whether the effect of a cognitive behavioral therapy (CBT) program on recidivism differ depending on probationer characteristics from a Bayesian perspective. Using Philadelphia CBT randomized controlled experiment data with a Bayesian hierarchical Gamma-Poisson model, the study compares average recidivism rates between the CBT and non-CBT groups, conditional on probationer characteristics. The Bayesian analysis showed that the effects of the CBT program were statistically meaningfully more evident for the high-risk probationers who were between 10-19 and 30-39 years old, who had more extensive prior experience on probation, and who had a higher ratio of high risk predictions. These results contradict the frequentist evaluation results that the CBT program had no statistically significant effect on recidivism.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGMENT	iv
ABSTRACT.....	vi
TABLE OF CONTENTS	viii
LIST OF TABLES.....	xi
LIST OF FIGURES.....	xiii
CHAPTER 1. LINKING THE ASSIMILATION OF IMMIGRANT YOUTH TO DELINQUENT RISK FACTORS.....	1
1.1. Introduction.....	2
1.2. Literature Review.....	4
1.2.1. Definition, Measure, and Theory of "Assimilation"	4
1.2.2. Studies on Immigrant Assimilation Effects	7
1.3. Method	12
1.3.1. Research Questions	12
1.3.2. Data Sources	12
1.3.3. Variables	14
1.3.4. Statistical Modeling	19
1.4. Results.....	25
1.4.1. Descriptive Statistics.....	25
1.4.2. Effect of Assimilation on Delinquency Risk Factors.....	28
1.5. Robustness Check	38
1.5.1. Bounding Exercise: Multiple Imputation Analysis of Missing Data	38

1.5.2. Revising Generation and Linguistic Assimilation Measures with Those Native-Born	42
1.6. Discussion and Conclusion	46
APPENDIX 1-A. Survey Question Item Statements	51
APPENDIX 1-B. Corresponding OLS Regression Results	54
CHAPTER 2. THE EFFECT OF THE INCREASED ALCOHOL AVAILABILITY ON CRIME: A QUASI-EXPERIMENTAL ANALYSIS OF THE REPEAL OF THE SUNDAY OFF-PREMISE LIQUOR SALES BAN	57
2.1. Introduction.....	58
2.2. Background and Literature Review	61
2.2.1. Pennsylvania "Blue Law" Repeal	61
2.2.2. Literature Review.....	63
2.2.3. Mechanisms: How Off-Premise Alcohol Sales are Connected to Crime.....	69
2.3. Study Design.....	72
2.3.1. Data Sources	72
2.3.2. Samples and Variables	73
2.3.3. Statistical Analysis Models.....	78
2.4. Results.....	83
2.4.1. Descriptive Statistics.....	83
2.4.2. Difference-in-Difference and Difference-in-Difference-in-Difference Results.....	89
2.4.3. The Geographical Displacement/Attraction Effect.....	91
2.4.4. Neighborhood Socioeconomic Status (SES) and the Displacement or Attraction Effect	95
2.5. Robustness Check	99
2.6. Discussion and Conclusion	103
APPENDIX 2-A. Variations in Numbers (#) of the W&S Stores in Philadelphia	109
APPENDIX 2-B. Real World Observations about W&S Stores in Philadelphia, PA	110
APPENDIX 2-C. DDD Results for the Full Set of the 94 W&S Stores.....	114

CHAPTER 3. A BAYESIAN ANALYSIS OF THE PHILADELPHIA PROBATION EXPERIMENT	119
3.1. Introduction.....	120
3.2. Literature Review.....	122
3.3. Data: Philadelphia Probation Experiment.....	125
3.4. Statistical Modeling	130
3.4.1. Research Questions	130
3.4.2. Bayesian Analysis in Criminology	130
3.4.3. Hierarchical Poisson Model	133
3.4.4. Prior	134
3.4.5. Posterior via the Monte Carlo grid sampling Method.....	135
3.4.6. Bayesian Approach for a Program Evaluation.....	138
3.4.7. Alternative: Frequentist Approach with Null Hypothesis Significance Test (NHST).....	140
3.5. Results.....	141
3.6. Robustness Test: Different Priors	152
3.7. Discussion and Conclusion	154
APPENDIX 3-A. An Exploratory Simulation for a Treatment-on-the-Treated Effect (ETT) Estimation	161
REFERENCES	165

LIST OF TABLES

Table 1-1. Descriptive Statistics of Dependent, Independent, Control Variables in the Current Rand Survey Data	26
Table 1-2. Logistic Regression: Outcome of Delinquent Peer Fraternization.....	32
Table 1-3. Logistic Regression: Outcome of Low Attachment to Family and School....	33
Table 1-4. Logistic Regression: Outcome of Positive Attitudes toward Delinquency	34
Table 1-5. Doubly Robust Estimations: Outcome of Three Delinquent Risk Factors.....	36
Table 1-6. Logistic Regression Odds Ratios Using 500 Hypothetical Missing-Imputed Data Sets.....	41
Table 1-7. Robustness Check: Logistic Regression Outcomes with Revised Assimilation Measures.....	44
Table 1-8. Robustness Check: DR Estimation Outcomes with Revised Assimilation Measures.....	45
Table 1-9. OLS Regression: Outcome of Delinquent Peer Fraternization	54
Table 1-10. OLS Regression: Outcome of High Attachment to Family and School.....	55
Table 1-11. OLS Regression: Outcome of Negative Attitudes toward Delinquency	56
Table 2-1. Types and Numbers of Crime Incidents, According to Days of the Week	84
Table 2-2. Average Daily Crime Incidents Occurring within the 1/8 mile radius areas of the 31 W&S stores.....	86
Table 2-3. Poisson Regression Average Marginal Effects in the DD and DDD Models	91
Table 2-4. Geographical Displacement/Attraction Effect: DDD Estimates for Crime Incidents Occurring Within Extended Radii of the W&S Stores (N=153,390)	94
Table 2-5. Geographical Displacement/Attraction Patterns for Low vs. High SES Neighborhoods	98
Table 2-6. Robustness Check: Alternative DDD Estimates for Crime Incidents Occurring Within 1/8 Mile and Between 1/4 and 1/2 Mile Radii of the W&S Stores ...	102

Table 2-7. Permutation Tests P-values for the Geographical Displacement Effect: DDD Estimates for Crime Incidents Occurring within Extended radii of the W&S Stores	103
Table 2-8. Variations in Numbers of the W&S Stores in Philadelphia	109
Table 2-9. Geographical Displacement/Attraction Effect: DDD Estimates for Crime Incidents Occurring within Extended Radii of the Full 94 W&S Stores (N=322,499) (25 Treatment and 69 Control Groups)	117
Table 2-10. Geographical Displacement/Attraction Patterns of the Full 94 W&S Stores for Low vs. High SES Neighborhoods	118
Table 3-1. Summary of the Outcome, "Any Violation Committed Within One Year".	127
Table 3-2. Proportions of Probationer Characteristic Categories	129
Table 3-3. Mean and Intervals of λ Levels and λ Contrast for All Treatment and Control Groups	144
Table 3-4. Mean and Intervals of λ Levels and λ Contrast, According to Age Subgroups	147
Table 3-5. Mean and Intervals of λ Levels and λ Contrast, According to Race Subgroups	149
Table 3-6. Mean and Intervals of λ Levels and λ Contrast, According to Prior Probation Experience	150
Table 3-7. Mean and Intervals of λ Levels and λ Contrast, According to Ratios of "High" Vote Subgroups	151
Table 3-8. Robustness Test: Estimates of Contrasts with Different Priors.....	153
Table 3-9. Mean and Intervals of λ Levels for All Control, All Treatment, and Treatment Attended Groups.....	162
Table 3-10. Distribution of 1000 simulated 95% Bayesian and Frequentist Intervals ..	164

LIST OF FIGURES

Figure 1-1. Conceptual Model between Assimilation, Delinquent Risk Factors, and Covariates	19
Figure 2-1. The Locations of the 6 Treatment and 25 Control Group W&S stores in Philadelphia	75
Figure 2-2. Average Yearly Total Crime Incidents Occurring within the 1/8 mile radius areas of the 31 W&S Stores	88
Figure 2-3. Average Marginal Effect Changes of Total Crime Incidents for Various Radii Areas	96
Figure 3-1. Contours of Unnormalized Joint Densities of (α, β) for the Control Group	144
Figure 3-2. Conditional Distribution of αY for the Control Group (N=447).....	145
Figure 3-3. Conditional Distribution of $\beta Y, \alpha$ for the Control Group (N=447).....	145

Chapter 1 . LINKING THE ASSIMILATION OF IMMIGRANT YOUTH TO DELINQUENT RISK FACTORS

Abstract

Despite increased academic interest in the relationship between immigrant assimilation and crime, relatively little is known about whether and how assimilation levels of immigrants are connected to the criminal and delinquent behaviors. The current study addresses this question by examining associations between a variety of assimilation measures and delinquent risk factors that may lead to delinquent behaviors, using a survey data from Los Angeles County, CA. The current paper employs a doubly robust estimation method that provides a protection against model misspecification, in addition to a conventional logistic regression method. The results show that immigrant youth with a low level of assimilation were more likely to have a high level of attachment to family and school than those with a high level of assimilation. However, due to small effect sizes and inconsistency of the statistical significances, the associations were less evident for the delinquent peer fraternization and positive attitudes toward delinquency outcomes. These results suggest that assimilation may be a weak predictor for delinquent risk factor development.

1.1. Introduction

Compared to the European immigrants of the early 20th century, the immigrant stream of the late 20th and early 21st centuries demonstrated different demographic and behavioral patterns. These "new" immigrants were foreign-born people and their children who migrated to the U.S. since the passage of the Immigration and Nationality Act Amendments in 1965. Most of them were from Latin America and Asia, comprising 51.7% and 26.4%, respectively, of all immigrants counted for the 2000 census (Gibson and Jung, 2006).

The new immigrants are characterized by the fact that despite a general lack of economic resources, they show noticeably positive behavioral patterns, especially in education and public health outcomes (e.g., Singh and Siahpush, 2002; Zhou and Bankston, 2006). This paradoxical phenomenon that worse living conditions are associated with better outcomes for the Hispanic/Latino immigrants is now referred to as the "Hispanic paradox" (Palloni and Arias, 2004; Scribner, 1996) or the "Latino Paradox" (Sampson, 2006 and 2008).

Recent empirical crime studies also provide consistent evidence that the new immigrants are less or at least no more prone to crime than their native-born counterparts (e.g. MacDonald and Saunders, 2012; Ousey and Kubrin, 2009; Reid et al., 2005; Stowell et al., 2009). With the accumulation of such empirical results, increasing attention is paid to how the new immigrants are connected to a low level of crime (Bersani, Loughram, and Piquero, 2013). Among possible mechanisms, assimilation, which can be defined in this paper as the unidirectional adaptation of a minority group to a host society's

dominant cultural customs, is considered as one of the strong candidates (Gibson and Miller, 2010; Smokowski, Dvaid-Ferdon, and Stroupe, 2009).

Recent studies attempted to directly gauge the degrees of association between assimilation measures and criminal and delinquent outcomes (e.g., Bui and Thongniramol, 2005), or to measure a mediating effect of crime risk factors on the association between assimilation and crime (e.g., Gibson and Miller, 2010). In the latter cases, crime or delinquent risk factors, such as delinquent peer fraternization, attachment to family and school, and attitudes toward crime and delinquency, were derived from relevant criminological or sociological theories.

However, the extant studies have some room to be improved yet. First, they typically use different single measures of assimilation, leaving research conclusions dependent on individual contexts, and impeding cross-validation of these measures (Smokowski, et al., 2009). Second, a majority of the studies rely on standard regression modeling only. Controlling for confounding by including it directly in the model structure, standard regressions are relatively vulnerable to misspecifications (Austin, 2011; Rubin, 2001). Third, despite the frequent appearance of risk factors as mediators, little is known about the relationship between assimilation and the employed risk factors. The Baron and Kenny's (1986) mediation analysis frame used in the extant studies requires significant associations between assimilation and the risk factors, and between the risk factors and crime outcomes. However, compared to the establishment of the latter association, research on the former association is rare.

The current paper aims to fill these research gaps in part. Using large-sized (N=1,435) self-reported survey data collected from Los Angeles, CA, this paper explores associations between three different delinquent risk factors as the outcome variables and four different assimilation measures. This paper also employs, in addition to conventional logistic regression, a doubly robust (DR) estimation method that combines conventional regressions and propensity scores-based weighting to render an additional protection against misspecification. With these settings, this paper expects that a low level of youth assimilation is less likely to be associated with delinquent risk factor development.

The remainder of this paper is organized as follows. Section 2 provides background information and brief literature reviews. Section 3 describes the logistic regression and DR analysis methods used in this paper. Section 4 reports empirical results from the two types of analyses. Section 5 provides robustness checks, one of which is a multiple-imputation analysis of missing data that suggests some bounds of statistical results. The paper ends with a summary of the findings and brief discussions of the implications, limitations, and directions for future research.

1.2. Literature Review

1.2.1. Definition, Measure, and Theory of "Assimilation"

Despite its common usage, the term "assimilation" remains in dispute in academia. More than often the term is used interchangeably with the term "acculturation." These two terms are defined in two ways, depending on the influencing directions between immigrant and native-born mainstream groups. While a bidirectional definition stresses

mutual influences and accommodations between the two groups,¹ a unidirectional definition emphasizes a non-dominant group that takes on the cultural customs of a dominant group in a host society.² There have been some academic attempts to define these terms.

One of them is Smokowski et al.'s (2009) tentative classification, which refers to *assimilation* as the unidirectional adaptation of a minority group to a host society's dominant cultural customs, and *acculturation* as the bidirectional cultural changes between minority and dominant groups. Based on this Smokowski et al.'s suggestion, the current study uses the term, assimilation, which emphasizes the unidirectional influence. While the extent to which immigrant youth adopt the English language in their normal lives is the main interest of the current study, the extent to which their native-born counterparts adopt an ethnic language, culture, and norms is not.

In addition to the definition issue, there also exists no consensus on how to accurately measure assimilation. For the unidirectional assimilation, a variety of single measures have been suggested so far. Based on the extant empirical studies that will be introduced in the next section, the current paper employed four different assimilation measures that are commonly used in the immigrant-crime studies: immigrant status, generation status, English-language use degree, and a gap of English-use degree use between immigrant youth and parents.

¹ "[P]henomena which results when groups of individuals having different cultures come into continuous first hand contact with subsequent changes in the original culture patterns of either or both groups..." (Redfield, Linton, and Herskovits, 1936: 149).

² "[T]he differences and changes in values and behaviors that individuals make as they gradually adopt the cultural values of the dominant society..." (Smith and Guerra, 2006: 283).

There are two major theoretical explanations of how immigrants are assimilated into U.S. society—the "classical" and "segmented" assimilation theories. The former argues that successful assimilation depends mostly on whether immigrants can interact with the dominant primary groups in the host society (Gordon, 1964; Park, 1936a and 1936b).³ According to this classical theory, the new Hispanic and Asian immigrants, who tend to settle in their ethnic enclaves, are likely to remain socially isolated from the mainstream primary group in U.S. society, thus being unlikely to be fully assimilated and improving their social status (usually referred to as "downward assimilation").

In contrast, the latter theory emphasizes that the segments of society into which immigrants are assimilated determine their futures (Portes and Zhou, 1993; Zhou, 1997a and 1997b; Zhou and Bankston, 1998 and 2006). According to this segmented assimilation theory, despite living in disadvantaged neighborhoods where deviant subculture norms are prevalent, immigrant youth who adhere to their ethnic customs can be protected by their ethnic communities, who provide strong informal social controls and supports. On the other hand, immigrant youth who abandon the customs of their ethnic communities are likely to absorb deviant subculture norms from surrounding underclass neighborhoods, such as positive attitudes toward crime and delinquency, which likely leads to criminal and delinquent behaviors (Portes and Zhou, 1993).

³ The patterns of the classical assimilation theory may be one example of the macro-sociological theory of social structure. According to Blau (1977: 52), the social structure of social positions affects the social relations between groups. One of his assertions is that the increase in heterogeneity promotes intergroup relations. Under the classical assimilation theory, an increasing number of immigrants makes a society more heterogeneous and prompt the interaction between the immigrant group and the primary dominant group. The immigrant group can then be assimilated into the mainstream through this interaction with the primary group.

1.2.2. Studies on Immigrant Assimilation Effects

A substantial body of empirical research has investigated the relationship between immigrant assimilation and crime-related outcomes in the fields of criminology, sociology, and public health. However, there exists no universal consensus on how to measure assimilation accurately. Most studies use a variety of single assimilation measures, including immigrant status, generation status, and language use. In addition, studies use Hispanic/Latino immigrants as their subjects, while only few investigate Asian immigrants. For outcomes, violent offending and youth delinquent behaviors are commonly employed in the studies. It is also notable that most studies utilize the conventional regressions or structural equation modeling.

A majority of the empirical studies gauge assimilation by the extent to which immigrants are assimilated linguistically. Language use has several advantages as an assimilation measure. First, it is easily measured through survey questionnaires. Second, it measures degrees of immigrant assimilation, which simple dichotomous measures like immigrant or generation status cannot. Third, as Vega and Gil (1998) and Finch et al. (2000) emphasized, language is not only the "conceptual core" of the host society's culture; it also represents the entire life of the culture into which one is assimilated. However, there are also concerns that this linguistic assimilation measure's predictive power may be low. Morenoff and Astor (2006) suggest time of immigration arrival as an alternative measure of assimilation, arguing that this measure is more powerful than the linguistic assimilation measure in predicting violent behaviors of immigrant youth.

Empirical studies using the linguistic assimilation measure have produced mixed results. For Hispanic immigrant youth, a higher level of linguistic assimilation was commonly associated with a higher likelihood of exhibiting violent (Morenoff and Astor, 2006) or delinquent behaviors (Vega et al., 1993). In addition, Spanish-speaking Hispanic adolescents were more likely to experience victimization through dating violence (Sanderson et al., 2004), to be bullied (Yu et al., 2003), and to be fearful of weapon-associated victimization than their English-speaking counterparts (Brown and Benedict, 2004). On the other hand, Decker et al. (2007) and Silverman, Decker, and Raj (2007) report no significant association between language use and sexual assault and dating violence for Hispanic and Asian immigrant youth.

There are studies that measure assimilation by using immigrant generation status.⁴ Bui and Thongniramol (2005) report that second-generation immigrant youth had a higher likelihood than first-generation youth for delinquent behaviors. Morenoff and Astor (2006) report that all the first-, second-, and third-generation adolescents differed from each other, and that the later generation had a higher likelihood of involvement in violent behaviors. Buriel, Calzada, and Vasquez (1982) reported that the third generation youth showed higher rates of delinquent behavior involvement than the first and second generation youth.

⁴ Commonly the first generation refers to as those who were foreign-born and migrated to the U.S. themselves, and the second generation indicates those who were native-born in the U.S. but have at least one foreign-born parent. The third generation refers to those who both were native-born in the U.S., and have two native-born parents. Meanwhile, there are people who were born abroad with foreign-born parents but migrated to the U.S. themselves at very early ages and were raised in the U.S. for almost all their lives. They are nearly the same as the second generation and frequently named as the "1.5 generation" (Rumbaut, 2004).

It is worth noting that a number of studies suggest mechanisms that mediate the link between immigrant assimilation and crime and delinquent behaviors. A variety of criminal and delinquent risk factors—including social bond, delinquent peer fraternization, deviance attitudes, self-control, family conflict, and parenting—are suggested as the mediators. It is also notable that virtually all are based on Baron and Kenny's (1986) so-called "causal-steps" mediation approach to test whether these mediators are effective.⁵

Gibson and Miller (2010) provide comprehensive empirical results, investigating the effects of assimilation, measured by linguistic assimilation and generational status, on both criminal offenses and violent victimization for immigrant youth. In general, a high level of assimilation was associated with an increase in offenses and victimization. However, while generational status was consistently a significant predictor of all types of

⁵ Baron and Kenny (1986) suggested an approach in which a variable can be "causally" tested as a mediator. The approach has been explained by the following diagram, where Y is regressed on X and M may be a mediator of the effect of X on Y. They proposed the following four steps:

Step 1: Conduct a simple regression of Y on X for path c: $Y = \beta_0 + c \cdot X + \epsilon$

Step 2: Conduct a simple regression of M on X for path a: $M = \beta_0 + a \cdot X + \epsilon$

Step 3: Conduct a simple regression of Y on M for path b: $Y = \beta_0 + b \cdot M + \epsilon$

Step 4: Conduct a multiple regression of Y on X and M for path c': $Y = \beta_0 + c' \cdot X + \beta_2 \cdot M + \epsilon$



outcomes, linguistic assimilation was significantly associated only with frequency of offenses. In addition, among the mediators suggested, delinquent peer fraternization was the only consistently significant mediator, while low self-control and inefficient parenting were not.

Parenting and family/parent-adolescent conflict are also commonly used as criminal and delinquent risk factors for studies on Hispanic populations. Dinh et al. (2002) suggest parent involvement as a significant mediator of the association between assimilation—measured by immigrant status and language use—and proneness to problem behavior. Gonzales et al. (2006) report that family conflicts mediated the positive association between linguistic assimilation and adolescent disorders. Samaniego and Gonzales (1999) report that Mexican adolescents' assimilation—measured through a combination of language use and generational status—was positively correlated with delinquent behaviors, being mediated by family conflict, inconsistent parent discipline, and maternal monitoring. Smokowski and Bacallo (2006 and 2007) and Sommers, Fagan, and Baskin (1993) also report that familism and parent-adolescent conflict were significant mediators of the association between assimilation and violence.

MacDonald and Saunders (2012)⁶ report that immigrant status was statistically significantly associated with a decrease in victimization, suggesting social bond to family and school and deviance-attitude learning as significant mediators of the association. Unlike the other studies, they employed the DR methods, which combine both conventional regression and propensity score-weighting techniques.

⁶ MacDonald and Saunders (2012) used the same data employed in the current paper.

Qualitative studies also describe and suggest the mechanisms that link immigrant assimilation to delinquency. One notable example is Portes and Rumbaut (2001), who note that immigrant youth who were less committed to ethnic communities' values and less attached to their parents were more likely to be involved in delinquent behaviors. In addition, these youth were most likely to be delinquent when they were substantially assimilated into the American customs and distant from their ethnic norms, while their parents still adhered to the ethnic customs of their nations of origin, which is known as "dissonant acculturation" (Portes and Rumbaut, 1996 and 2001).

In sum, the literature review shows that the following issues need to be addressed further. First, assimilation is measured in a variety of ways, including immigrant status, generation status, linguistic assimilation degree, and family dissonant linguistic assimilation gap. The empirical results tend to be dependent on the types of measures. A cross-validation check of these different measures is viable. Second, despite suggestions of criminal and delinquent risk factors as mediators, including delinquent peer fraternization, low attachment to family and school, little is known about the association between assimilation and the risk factors suggested. Third, most of the extant studies rely on the conventional regression method only, having different control variables and different functional forms. The empirical results may have differed if the regression specifications were different.

1.3. Method

1.3.1. Research Questions

This paper empirically tests the following questions about the association between immigrant assimilation and the delinquent risk of immigrant youth.

Question 1. For immigrant youth, is a higher level of assimilation associated with a higher likelihood of delinquent risk factor development?

Question 2. Are the associations between assimilation and the outcomes consistently evident across different types of assimilation measures and estimation methods?

1.3.2. Data Sources

The individual- and household-level data used for this study were collected mainly as part of a business-improvement district (BID) self-reporting survey conducted by the Rand Corporation (MacDonald et al., 2009).⁷ The data consisted of two cross-sectional sets gathered from August 2008 to May 2009 and from September 2009 to March 2010, respectively. Due to the temporal closeness and equivalency of the questionnaires used, the two waves were considered as one cross-sectional survey. The final-version survey contained a total of 1,435 eligible households. In addition, neighborhood-level demographic and social variables of the census tracts in which the

⁷ I am grateful to Dr. John M. MacDonald for providing access to this Rand Survey data set.

households were located were extracted from the 2005-2009 American Community Survey five-year estimates.

For each household sampled, questions were asked of one parent and one of the children targeted to be ages 14-17 in English or Spanish. However, 53 parents neglected to check their immigrant or native statuses (3.7% missing). Of the remaining 1,382 household parents, 797 responded that they were foreign-born, thus being classified as "immigrant." Among the children in this group, 679 youth indicated whether they were born in the U.S. or abroad for generation status (14.8% missing).

However, it is worth noting that there is a portion of youth who responded that they spoke languages other than English at home, showing non-perfect language assimilation levels, 73 of which were in native-born households and 37 of which had no immigrant status information.⁸ The pool for language-related assimilation measures included these 110 abnormal cases in addition to the 797 immigrant household samples, generating a total of 907 eligible household samples. Note that as a result, the object youth group of the language-related assimilation measures were not equal to that of the youth group for the immigrant status and generation status measures in this study. Among the 907 youth, 682 youth answered the questions about their English-speaking

⁸ These 110 youth may have been in households whose not-surveyed parent was foreign-born, or they were third-generation immigrant youth whose parents were second generation, thus adhering to the ethnic customs of the origin countries. If this was the case, despite the nominal "native-born" classification, their households might be indeed immigrant households.

assimilation levels (24.8% missing), and both parents and children of 488 households answered these questions (46.1% missing).⁹

1.3.3. Variables¹⁰

Dependent variables included three delinquent risk factors: fraternization with delinquent peers, low attachment to family and school, and positive attitudes toward delinquency. The first two delinquent risk factors meshed with the variables were relatively common in other empirical studies. The last risk factor, positive attitudes toward delinquency, was constructed as the operational definition of the deviance subcultural norms of the segmented assimilation theory.

Related question items were combined to construct these dichotomous dependent variables.¹¹ When items were simply "Yes or No" dichotomous questions, dependent variable were assigned to 1 if any of the items was answered "Yes" and to 0 if all of the items were answered "No." "Delinquent Peer Fraternization," derived from nine items (Cronbach's $\alpha=0.71$) was made in this way. When items were measured on more than two scales, dependent variables were differently constructed. They were assigned to 1 if the item-averaged scale was over a certain threshold and to 0 if the scale was below the

⁹ The high missing proportions are another concern of this study, because, if the missing samples had different characteristics and behavioral patterns than those of the present samples, the currently reported analysis results might be biased. The multiple imputation analysis in the robustness check section provides a set of possible bounds of the current results.

¹⁰ See Appendix 1-A to check the item lists of each variable.

¹¹ Language-related independent variables are made dichotomous to be consistent in measurement with the other immigrant status and generation status variables. In addition, the DR estimation requires a dichotomous independent variable for which a propensity score can be calculated. See the Appendix 1-B for OLS statistical results with continuous language-related independent variables and continuous dependent variables.

threshold. "Low Attachment to Family and School," derived from eight items (Cronbach's $\alpha=0.70$) having five Likert scales, was assigned to 1 when the average scale was less than four ("Not at all," "Very little," or "Somewhat") and 0 if it was equal to or over four ("Quite a bit" or "Very much"). In addition, "Positive Attitudes toward Delinquency," derived from six items (Cronbach's $\alpha=0.66$) having four Likert scales, was assigned to 1 when the average scale was less than three ("Strongly agree" or "Somewhat agree") and to 0 if it was equal to or more than three ("Unsure" or "Strongly disagree").

Independent variables were four different assimilation measures. The first was immigrant status indicating whether the surveyed household parent was foreign-born (born outside the U.S. or one of its territories) or native-born (born in the U.S. or one of its territories). If the household parents were foreign-born, their children were likely to be exposed to the ethnic customs of the parents' original countries. Compared to native-born non-immigrant youth, these immigrant youth were less likely to be assimilated into the U.S. culture. In that sense, immigrant status can be a rough measure of assimilation. Being a dichotomous variable, this "Immigrant Status" variable was assigned to 1 when the surveyed parent was foreign-born and to 0 when the parent was native-born.

Another assimilation measure was the dichotomous "Generation Status." Being born outside the U.S. might be associated with a shorter residence in the U.S. and less time to be assimilated. In the current study, this variable distinguished between

immigrant youth of 1st generations who were born abroad (including 1.5th generations) and those of 2nd and 3rd generations who were born in the United States¹².

The third and main assimilation measure in this paper was the "Linguistic (English) Assimilation" level, which was derived from four items (Cronbach's $\alpha=0.63$) having five Likert scales. As noted in the literature review, language is considered as the core of the assimilation process (Vega and Gil, 1998; Finch et al., 2000). The dichotomous variable was then assigned to 1 if the average over the four item scales was equal to or less than 3.5, being "Low-assimilated," and 0 if the averaged scale was more than 3.5, being "High-assimilated"¹³

The final assimilation measure, "Dissonant Linguistic Assimilation Gap," was constructed based on the differences between immigrant youth's and their parents' linguistic assimilation measures in the same households. This assimilation measure was derived, as the operational definition of the concept of "dissonant assimilation," from Portes and Rumbaut (2001). The larger the dissonant linguistic assimilation gap between the immigrant youth and their parents are, the more likely the youth are involved in delinquent behaviors. Given that a majority of the current data samples were Hispanic (77.4% of the total) and that Hispanic immigrant parents are usually not fluent in English, a small dissonant gap likely corresponds to a low level of linguistic assimilation of immigrant youth, while a large gap does to a high level of linguistic assimilation of

¹² Due to data limitations, it was not possible to distinguish between the second and third generations in the current data set.

¹³ This 3.5 scale threshold may be controversial because the value was determined discretionally. However, it was impossible to obtain a universal consensus on which value could be the best threshold; and tweaking the threshold in a small range around 3.5 did not change the results dramatically, which renders supports for this threshold decision.

immigrant youth. When the linguistic assimilation scales of the youth were subtracted by those of their parents, the dissonant linguistic assimilation gap variable was assigned to 0 if the assimilation difference was higher than 1 ("Large Dissonant Assimilation Gap" = 0) and 1 if it was equal to or less than one ("Small Family Dissonant Assimilation Gap"=1).¹⁴

In addition, some background characteristics at the individual, household, and neighborhood levels were added as control variables. At the individual level, youth ages, Hispanic ethnicity, and genders were added.¹⁵ The youth ages were categorized into five scale groups, reflecting the original target age range from 14 to 17: scale 1 for ages equal to or younger than 14 (12%), scale 2 for age 15 (16%), scale 3 for age 16 (25%), scale 4 for age 17 (25%), and scale 5 for ages 18 or older (23%). However, the actual age distribution of the surveyed youth in the current data set was 9 to 23.¹⁶ Also, at the household level, a household socioeconomic status (SES) variable was constructed by combining three variables of parent education level, household income, and average monthly mortgage or rent, having a range from 0 to 6 scales.

¹⁴ One interesting feature about this dissonant assimilation measure was that around 8% (39 out of the 488 cases) reported that the parents' linguistic assimilation scales were higher than those of their children's, thus yielding minus values of the assimilation differences. One possible explanation could be that the surveyed immigrant parents exaggerated their degrees of language assimilation.

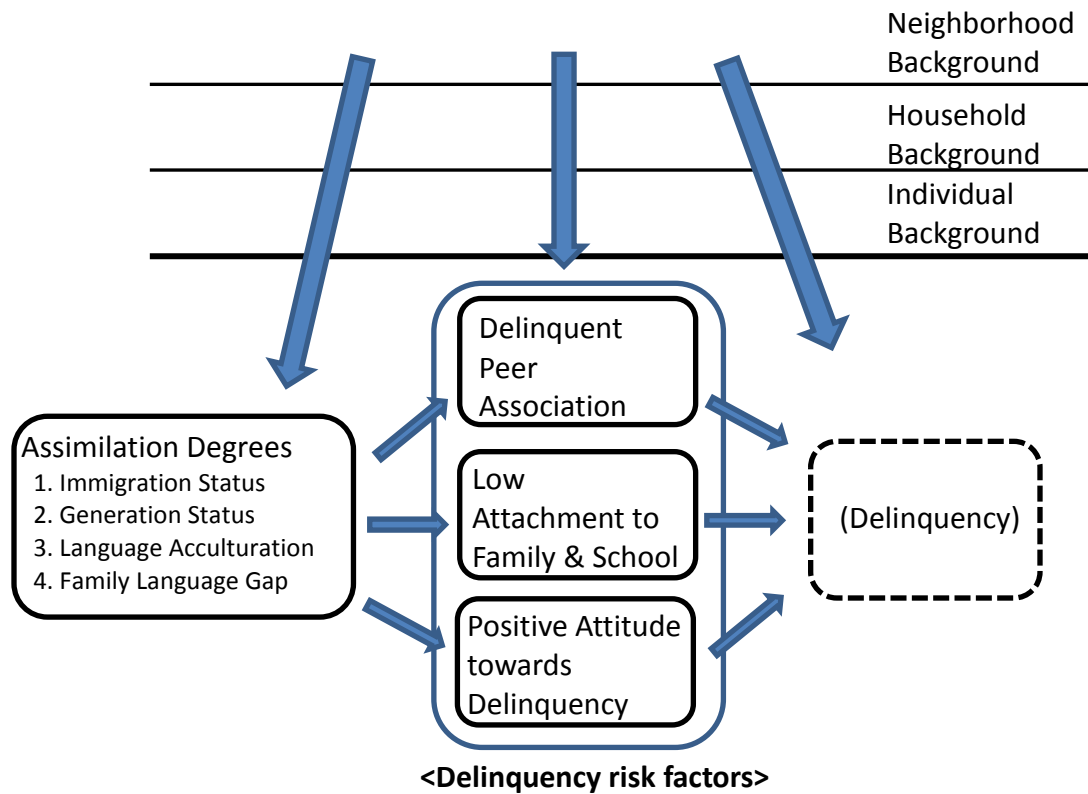
¹⁵ The races of the youth were not included as covariates in this analysis because the race classification in the survey was not clear and could have distorted the results. The survey results showed that about half the surveyed youth were white, with black (4.4%) or Asian (4%) comprising small segments. The remaining 40% of the surveyed youth marked themselves as being Hispanic or belonging to more than one race.

¹⁶ One reason for this out-of-target age range was that the first wave of the data, which was collected between August 2008 and May 2009, was actually the follow-up of another original RAND BID survey. Those who responded to the original survey at the targeted ages were followed up 18 months later and asked to complete a new survey, which was being currently used, at the older ages than 17. Another reason might be coding errors.

At the neighborhood level, three composite neighborhood variables were constructed, based on Sampson, Raudenbush, and Earls (1997). Two were derived from nine structural neighborhood variables of the 2005-2009 American Community Survey. "Concentrated Immigration" was constructed by combinations of percentages of Hispanic population and percentages of those foreign-born, while "Concentrated Disadvantaged" comprised combinations of six variables: percentages of the population ages 10 to 19, percentages of blacks, percentages of female-headed families, percentages of households whose incomes were below the poverty line, percentages of households that received food-stamp assistance in the past 12 months, and percentages of unemployed. The other "Collective Efficacy" variable, which stands for the neighborhood social cohesion and informal control, was constructed from the parents' 15 items on the neighborhood cohesion in the RAND survey (Cronbach's $\alpha=0.87$).

Figure 1.1 presents a conceptual model in which the all variables of assimilation measures, delinquent risk factors, and covariates are assumed to be associated. Note that, reflecting the lack of direct measures of youth's delinquent behaviors in the current study, the outcome of delinquent behaviors was depicted by a dotted perimeter in the figure. Note also that the arrows in this conceptual model figure did not mean a causal relationship; the current study was rather exploratory, and the arrows merely indicated the investigation directions of the current study.

Figure 1-1. Conceptual Model between Assimilation, Delinquent Risk Factors, and Covariates



1.3.4. Statistical Modeling

A primary statistical method in this paper was a logistic regression.¹⁷ Note that the current regression model was not a hierarchical model, despite the three different unit layers of individual, household, and neighborhood levels in the data set. There were too few observations for each of household (only one youth and one parent) and neighborhood (on average fewer than 4 households). Instead, the robust standard error (the Huber-White Sandwich Estimator) was used to control for clustering effects at the household and neighborhood level, in place of a hierarchical model.

¹⁷ OLS regressions were also conducted. Their results are reported in Appendix 1-B.

$$\text{Logit}(Y) = \beta_0 + \beta_1 \cdot Z + \underline{\beta_2} \cdot \underline{X} + \varepsilon$$

Where Y = delinquency risk factors of youth

Z = a youth assimilation measure:

- (1) immigrant status, or
- (2) generation status, or
- (3) youth linguistic assimilation level, or
- (4) family dissonant linguistic assimilation gap

\underline{X} = youth, family, and neighborhood characteristics vector, including:

- youth age
- youth gender
- youth Hispanic
- household SES (from parent education level, household income, and average monthly mortgage or rent)
- Concentrated Disadvantage (from % black, % population aged 10~19, % female-headed family, % households that received food-stamp assistance in the past 12 months, % unemployed, and % households whose income is below the poverty line),
- Concentrated Immigration (from %Hispanic, and %foreign-born),
- Collective Efficacy

While the conventional regression methods are valuable tools for estimating an average treatment effect, they have limitations that need to be addressed. One of them is that regression results do not necessarily represent a population-level average effect similar to what can be obtained from randomized controlled experiments (Austin, 2011). In the potential outcomes model frame, subject-specific individual treatment effects are not necessarily equal to population-level treatment effects for observational data.¹⁸ The former is estimated by comparing the outcome risk for a subject who is exposed to treatment to what it would have been if the subject had not been exposed to treatment, or by comparing the outcome risk for a subject who is not exposed to treatment to what it

¹⁸ In epidemiology, the former is called as a conditional average treatment effect, while the latter is as a marginal average treatment effect (Austin, 2011).

would have been if the subject had been exposed to treatment. The latter is estimated by comparing a population-level average effect on outcome if the entire population had been exposed to treatment, to population-level average effect on outcome if the entire population had been not exposed to treatment (Clayton and Hills, 1993). While individual-level average effects can be estimated by conventional regression methods that compare subjects having similar covariate values, population-level average effects cannot be, although these two effects may be collapsible when there is no confounding (Greenland, 1987).

Another limitation is that conventional regressions are relatively vulnerable to misspecification (Joffe et al., 2004). Regression results depend on the structural part of a statistical model, which consists of variables included for controlling for confounding, based on a variety of assumptions such as a functional form assumption. To control for confounding, mere nuisance variables may be included in the model structure (Joffe et al., 2004). The omitted variable bias is another threat to the modeling that cannot be fully addressed. In addition, because it is relatively easy to manipulate structures, there always exists a temptation for researchers to continually over specify the model until they can obtain the desired results (Austin, 2011; Rubin, 2001). Unfortunately, whether a regression model is specified properly is not statistically verifiable.

Compared to conventional regressions, propensity score methods have relative strength for the aforementioned issues (Austin, 2011; Joffe et al., 2004). The propensity-score methods were suggested as one way to obtain the population-level average effect with observational data (Rosenbaum and Rubin, 1983). For example, as one of the

methods, the inverse probability of treatment weighting (IPTW) facilitates the construction of a pseudo-population that has no confounding influences from the given covariates. A comparison of the treatment and control groups in the covariate-balanced pseudo-population is supposed to yield a population-level average treatment effect. Also, the propensity score method is relatively easy in model specification (Austin, 2011). By separating the model selection from the confounding controlling, the propensity score method enables an output model to be specified without including nuisance variables (Joffe et al., 2004). Also, this separation prevents overspecification temptation.

However, despite its relative easiness in model specification compared to conventional regressions, this propensity score-based method is not totally free from specification errors. In the first stage where the propensity scores are predicted, the propensity score method counts only observable variables, leaving unobservable characteristics unexplained.¹⁹ If propensity scores are not truly representative of necessary covariates that model the treatment selection, average effects estimated in a propensity score model are biased (Lunceford and Davidian, 2004).

Given various possible biases due to misspecification, this paper attempts to estimate a population-level average treatment effect from the doubly robust (DR) estimation model (Bang and Robins, 2005; Emsley et al., 2008; Lunceford and Davidian,

¹⁹ In this sense, the propensity score method may have a limited ability to address selection bias. Treatment selection tends to be related to subject characteristics (Austin, 2011). If determinants of treatment selection are fully observable as covariates and considered in predicting propensity scores, then the propensity methods may address the selection bias by having the populations that have balanced covariates (Ye and Kaskutas, 2009). However, it appears to be a rare case that all the selection determinants are observed in the real world. Therefore, by balancing observable selection determinants across treatment and control populations only, the propensity-score method appears to only partially address the selection bias.

2004). This model combines conventional regression modeling and propensity score weighting into one model. Confounding effects are adjusted for in two ways: by modeling the effects of the confounders on outcomes explicitly (in the outcome model) and by weighting observations based on the inverses of the predicted probabilities of treatment (in the exposure model). While this DR method allows the advantage of estimating a population-level average treatment effect, it provides an additional protection advantage against misspecification; if either the outcome model or the exposure model is correctly specified, the DR estimate returns unbiased results.²⁰

As a caveat of this model, there may be a loss of precision (Lunceford and Davidian, 2004). If the propensity score modeling in the exposure model is correct, the DR estimator will have a smaller variance than the simple inverse-propensity weighted estimator in large samples. Meanwhile, if the regression modeling in the outcome model

²⁰ The mathematical notation for the double robust estimator is as follows (Lunceford and Davidian, 2004):

$$\hat{\Delta}_{dr} = \frac{1}{n} \sum_{i=1}^n \left[\frac{Z_i Y_i}{e(X_i, \hat{\alpha})} - \frac{\{Z_i - e(X_i, \hat{\alpha})\}}{e(X_i, \hat{\alpha})} m_1(X_i, \hat{\beta}_1) \right] - \frac{1}{n} \sum_{i=1}^n \left[\frac{(1-Z_i) Y_i}{1 - e(X_i, \hat{\alpha})} - \frac{\{Z_i - e(X_i, \hat{\alpha})\}}{1 - e(X_i, \hat{\alpha})} m_0(X_i, \hat{\beta}_0) \right]$$

$$= \hat{\mu}_1, dr - \hat{\mu}_0, dr$$

where $\hat{\mu}_1, dr$ and $\hat{\mu}_0, dr$ are the treatment and control group effects in the DR model, respectively, $e(X, \alpha)$ is a postulated model for the true propensity score $e(X) = E(Z|X)$, and $m_1(X, \beta_1)$ and $m_0(X, \beta_0)$ are postulated models for the true regression $E(Y|Z=1, X)$ and $E(Y|Z=0, X)$, respectively.

Substituting α, β for the estimated quantities, where α is an outcome model coefficient vector and β is an exposure model coefficient vector,

$$E\left[\frac{Z Y_1}{e(X, \alpha)} - \frac{\{Z - e(X, \alpha)\}}{e(X, \alpha)} m_1(X, \beta)\right] - E\left[\frac{(1-Z) Y_0}{1 - e(X, \alpha)} - \frac{\{Z - e(X, \alpha)\}}{1 - e(X, \alpha)} m_0(X, \beta)\right]$$

$$= E(Y_1) + E\left[\frac{\{Z - e(X, \alpha)\}}{e(X, \alpha)} \{Y_1 - m_1(X, \beta)\}\right] - E(Y_0) - E\left[\frac{\{(1-Z) - (1 - e(X, \alpha))\}}{1 - e(X, \alpha)} \{Y_0 - m_0(X, \beta)\}\right]$$

If either a postulated propensity score model $e(X, \alpha) \leftarrow e(X) = E(Z|X) = E(Z|Y, X)$ (by no unmeasured confounders) or a postulated regression model $m(X, \beta) \leftarrow E(Y|Z, X) = E(Y|X)$ (by no unmeasured confounders) is correct, then $E\left[\frac{\{Z - e(X, \alpha)\}}{e(X, \alpha)} \{Y_1 - m_1(X, \beta)\}\right] = 0$, and $E\left[\frac{\{(1-Z) - (1 - e(X, \alpha))\}}{1 - e(X, \alpha)} \{Y_0 - m_0(X, \beta)\}\right] = 0$, so that $\Delta_{dr} = E(Y_1) - E(Y_0)$, which is an unbiased estimate for an average effect.

is correct, the estimator may have a larger variance than the simple regression estimator in large samples. However, because we do not know whether or not a model is ultimately properly specified, the protection function of the DR estimation is warranted.

In the current paper, a DR estimation model was set, as a complementary secondary model to the conventional regression. Identical covariates were included in both the outcome and exposure models. However, note that because two of the four assimilation measures—immigrant status and generation status—had been predetermined to be independent of individual, household, and neighborhood factors for immigrant youth, these two assimilation measures were excluded from the DR analysis (Austin, 2011). Therefore, the DR estimations were conducted only for the two language-related assimilation measure cases.

- **Outcome Model** (fitted by a logistic regression modeling)
: Dichotomous Delinquent-Risk Factors
= Dichotomous Youth Assimilation or Dissonant Assimilation Gap
+ Youth Characteristics (*age, gender, and Hispanic*) + Household SES
+ Neighborhood Concentrated Disadvantage (*% black, % population ages 10~19, % female-headed family, % households that received food-stamp assistance in the past 12 months, % unemployed, and % households whose income is below the poverty line*)
+ Neighborhood Concentrated Immigration (*% Hispanic, %foreign-born*)
+ Collective Efficacy Index
- **Exposure Model** (fitted by a propensity score weighting)
: Dichotomous Youth Assimilation or Dissonant Assimilation Gap

= Youth Characteristics (*age, gender, and Hispanic*) + Household SES
 + Neighborhood Concentrated Disadvantage (*% black, % population ages 10~19, % female-headed family, % households that received food-stamp assistance in the past 12 months, % unemployed, and % households whose income is below the poverty line*)
 + Neighborhood Concentrated Immigration (*%Hispanic, %foreign-born*)
 + Collective Efficacy Index

1.4. Results

1.4.1. Descriptive Statistics

Table 1.1 shows descriptive statistics of the current data samples. The percentages represented portions of those who answered "1=yes" among all those who answered for given dichotomous questions. For the dependent variables, there was no statistically significant difference in sample proportions across assimilation levels, except for two cases—immigrant youth were more likely to have positive attitudes toward delinquency than native-born youth (38% vs. 30%), and youth with small family dissonant assimilation gaps were less likely to have low attachment to family and school than youth with large family dissonant assimilation gaps (16% vs. 23%). For the independent variables, there were consistently statistically significant differences across assimilation levels for the four assimilation measures.

Table 1-1. Descriptive Statistics of Dependent, Independent, Control Variables in the Current Rand Survey Data

		All Youth	Immigrant vs. Native-born Household (total N= 1,435)		1 st vs. 2 nd /3 rd Immigrant Generation (total N= 797)		Low vs. High Linguistic Assimilation (total N= 907)		Small vs. Large Dissonant Linguistic Assimilation Gap (total N= 907)	
			Immi-grant	Native-born ¹	1 st G.	2 nd /3 rd G.	Low-Assim. 1, 2	High-Assim. 1, 2	Small-Gap ^{1, 2}	Large-Gap ^{1, 2}
Dependent Variable	Delinquent Peer Fraternization (dummy: 1=Yes)	43% (N=1,435)	41% (N=797)	45% (N=585)	50% (N=135)	52% (N=544)	50% (N=382)	54% (N=300)	52% (N=275)	50% (N=213)
	Low Attachment to Family & School (dummy: 1=Yes)	22% (N=1,124)	21% (N=632)	24% (N=441)	19% (N=135)	22% (N=543)	18% (N=382)	24% (N=300)	16% (N=275)	23%* (N=213)
	Positive Attitude Toward Delinquency (dummy: 1=Yes)	35% (N=1,123)	38% (N=632)	30%* (N=440)	39% (N=135)	38% (N=543)	38% (N=381)	40% (N=300)	37% (N=275)	37% (N=213)
Independent Variable	Immigrant Household (dummy: 1=Yes)	58% (N=1,382)	-	-	-	-	94% ³ (N=359)	82%* ³ (N=286)	87% ³ (N=274)	98%* ³ (N=211)
	1 st Immigrant Generation (dummy: 1= Yes)	80% (N=629)	80% (N=629)	-	-	-	24% (N=335)	17%* (N=235)	25% (N=237)	14%* (N=205)
	Youth Language Assimilation Index (mean scale: 1~ 5)	3.54 ^{1, 2} (N=682)	3.52 (N=572)	-	3.38 (N=129)	3.55* (N=478)	3.19 (N=382)	4.00* (N=300)	3.47 (N=275)	3.63* (N=213)
	Youth Low Assimilation (dummy: 1=Yes)	56% ^{1, 2} (N=682)	59% (N=572)	-	67% (N=129)	57%* (N=478)	-	-	62% (N=275)	49%* (N=213)
	Dissonant Assimilation Gap Index (mean scale: -4 ~ 4)	1.02 ^{1, 2} (N=488)	1.09 (N=443)	-	0.86 (N=88)	1.15* (N=357)	0.85 (N=277)	1.23* (N=211)	0.42 (N=275)	1.78* (N=213)
	Small Family Dissonant Assim. Gap (dummy: 1=Yes)	56% ^{1, 2} (N=488)	54% (N=443)	-	68% (N=88)	50%* (N=357)	62% (N=277)	49%* (N=211)	-	-

Control Variable	Youth Age (mean scale: 1~5)	3.32 (N=1,088)	3.29 (N=610)	3.31 (N=432)	3.38 (N=132)	3.55 (N=521)	3.32 (N=367)	3.34 (N=288)	3.19 (N=259)	3.05 (N=205)
	Youth Gender (dummy: 1=male)	52% (N=1,125)	50% (N=633)	55% (N=441)	55% (N=135)	49% (N=544)	48% (N=382)	56%* (N=300)	51% (N=275)	53% (N=213)
	Youth Hispanic (dummy: 1=Hispanic)	61% (N=1,124)	85% (N=632)	25%* (N=441)	82% (N=135)	85% (N=544)	95% (N=382)	78%* (N=300)	82% (N=275)	95%* (N=213)
	Household SES (mean scale: 1~6)	3.13 (N=1,375)	2.42 (N=790)	4.11* (N=581)	2.34 (N=129)	2.35 (N=500)	2.10 (N=356)	2.71* (N=288)	2.71 (N=274)	1.90* (N=211)
	Concentrated Disadvantage (%)	19% (N=1,434)	21% (N=797)	15%* (N=585)	21% (N=135)	22% (N=543)	22% (N=382)	20%* (N=299)	21% (N=275)	23%* (N=213)
	Concentrated Immigration (%)	46% (N=1,434)	54% (N=797)	35%* (N=585)	55% (N=135)	55% (N=543)	58% (N=382)	51%* (N=299)	54% (N=275)	59%* (N=213)
	Neighborhood Collective Efficacy (mean scale: 1~4)	2.87 (N=1,394)	2.76 (N=797)	3.03* (N=585)	2.74 (N=129)	2.77 (N=510)	2.75 (N=363)	2.80 (N=291)	2.76 (N=275)	2.78 (N=213)

*: The p-value < 0.05, implying that two compared groups are statistically significantly different for the given characteristic at the 95% confidence level.

¹: 73 youth had native-born parents but answered the language-assimilation questions, including that they spoke languages other than English at home. These youth were counted as native-born. However, due to their incomplete linguistic assimilation levels, they were included in the linguistic assimilation measure calculations.

²: 37 youth answered the language-assimilation questions but their parents did not answer whether they themselves were foreign-born. These youth were not included in the immigrant-status and generation-status measure calculations. However, due to their incomplete linguistic assimilation levels, they were included in the linguistic assimilation measure calculations.

³: These immigrant household percentages were not 100% because the 73 native-born and 37 non-immigrant non-native youth were included.

For the control variables, there were significant differences across levels for the three assimilation measures, with the exception of generation status. Immigrant youth differed from native-born youth in the proportions who were Hispanic, levels of household SES, degrees of neighborhood-concentrated disadvantage, degrees of neighborhood-concentrated immigration, and levels of neighborhood collective efficacy. Youth groups who had low and high levels of linguistic assimilation differed in proportions of female youth, proportions of youth who were Hispanic, levels of household SES, degrees of neighborhood-concentrated disadvantage, and degrees of neighborhood-concentrated immigration. At the same time, youth groups who had small and large dissonant linguistic assimilation gaps with their parents differed in the proportions who were Hispanic, levels of household SES, degrees of neighborhood-concentrated disadvantage, and degrees of neighborhood-concentrated immigration. In contrast, 1st and 2nd/3rd generations of immigrant youth had no significant differences for the all control variables.

1.4.2. Effect of Assimilation on Delinquency Risk Factors

Table 1.2 provides logistic regression results with the delinquent peer fraternization outcome and four assimilation measures. All the values reported in the table cells were in odds ratios. In Model 1, the dichotomous delinquent peer fraternization status was regressed on the immigrant status. The odds ratio for this variable was 0.62, being statistically significant at the two-tailed 0.05 level. The odds of fraternizing with delinquent friends were 38% less for immigrant youth than for native-

born youth, thus implying that immigrant youth were less likely to fraternize with delinquent peers than native-born youth.

Model 2 in Table 1.2 shows the regression result for generation status, having an odds ratio coefficient of 0.89. In Model 3, the odds ratio for linguistic assimilation status was shown to be 0.90. While the odds ratios for the other assimilation measures in Model 1 to 3 were less than 1, the odds ratio for the assimilation measure of dissonant linguistic assimilation gap status in Model 4 was greater than 1, being 1.09. However, because the coefficients of Models 2, 3, and 4 were statistically insignificant, no statistical conclusion can be made from these results.

Most individual covariates (except for Hispanic with dissonant linguistic assimilation measure), concentrated disadvantage, and collective efficacy) were consistently statistically significant predictors of delinquent peer fraternization across the assimilation measures. When youth were older, male, and Hispanic, they were more likely to fraternize with delinquent friends. At the same time, those living in disadvantaged neighborhoods and in neighborhoods having higher levels of collective efficacy were less likely to fraternize with delinquent friends. On the other hand, household SES and concentrated immigration covariates were not statistically significant in predicting delinquent peer fraternization.

Table 1.3 provides logistic regression results that examine whether youth had a low level of attachment to family and school. In Model 1, the odds ratio for the immigrant status was 0.63, being statistically significant at the two-tailed 0.05 level. This 0.63 odds ratio indicates that holding the other values equal, the odds of having a low

level of attachment to family and school over having a high level of attachment to family and school were about 37% less for immigrant youth than for native-born youth. In other brief words, immigrant youth are more likely to feel attached to family and school than native-born youth.²¹

Both language-related assimilation measures were also statistically significant predictors of youth's attachment levels at the two-tailed 0.05 level. For the measure of linguistic assimilation status (Model 3), the odds ratio coefficient was 0.61. This means that, holding the other values equal, the odds of having a low level of attachment to family and school over having a high level of attachment to family and school were about 39% less for youth who had a low level of linguistic assimilation than for youth who had a high level of linguistic assimilation. Also, the family dissonant linguistic assimilation gap status measure had an odds ratio coefficient of 0.60 (Model 4), indicating that, holding the other values equal, for youth having small linguistic assimilation gaps with their parents, the odds of having a low level of attachment to family and school over having a high level of attachment to family and school were about 40% less than for youth having large gaps with their parents. In other words, the less language-assimilated

²¹ The current attachment outcome was constructed by both attachment to family and attachment to school measures. The current attachment outcome variable could be broken down into two separate attachment outcomes — low levels of attachment to family and to school. However, the odds ratio results for these two new outcomes were in general supportive of the original attachment outcome results. Above all, youth having a low level of linguistic assimilation were likely to feel more attached to family (odds ratio coefficient: 0.36 at the 1% level) and to school (odds ratio coefficient: 0.66 at the 5% level) than those with a high level of linguistic assimilation. For the other assimilation measures, there were some differences in statistical significances. Youth having similar linguistic levels with their parents are likely to feel more attached to family only, compared to those having different linguistic levels with their parents (odds ratio coefficient: 0.45 at the 5% level). Meanwhile, first immigrant youth are likely to feel more attached to school only, compared to second or third generation immigrant youth (odds ratio coefficient: 0.67 at the 10% level). An immigrant status is no longer significant predictor of the attachment outcomes. The full statistical results are available on request.

youth were, or the smaller dissonant English-assimilation gaps youth had with their parents, the more likely they were to feel attached to their family and school.

However, the assimilation measure of generation status (in Model 2) and virtually all the control variables were poor predictors of the youth's levels of attachment to family and school. The odds ratio coefficients were statistically insignificant, even at the two-tailed 0.1 level, except for some of the age-of-youth and collective efficacy variables.

Table 1.4 reports the logistic regression odds ratios for the outcome of positive attitudes toward delinquency. Interestingly, all four assimilation measures consistently had statistically insignificant coefficients, even at the two-tailed 0.1 level, indicating that all the assimilation measures were poor predictors of the outcome. Instead, the household SES, neighborhood concentrated disadvantage, and neighborhood collective efficacy control variables had some statistically significant odds ratio coefficients, but only as being inconsistent across all four models.

Table 1.5 reports the doubly robust (DR) odds ratio estimates for the two language-related assimilation measures. The estimate results are closer to 1, compared to the previous logistic regression results. In this DR estimation, the treatment groups consist of youth who had low levels of linguistic assimilation and youth who had small dissonant gaps of linguistic assimilation with their parents, compared to those who had high levels of linguistic assimilation and those who had large dissonant gaps, respectively. They were weighted by the propensity scores to have balanced covariates with less than 10% standardized mean differences and less than 0.1 standard deviation ratios. Also, the

DR estimates are population-level average effects, which differ from the subject-level average effects of regression-model estimates.

Table 1-2. Logistic Regression: Outcome of Delinquent Peer Fraternization

	DV: Delinquent Peer Association (1=Yes)			
	Model 1: Immigrant Status (N=1036)	Model 2: Generation Status (N=607)	Model 3: Youth Linguistic Assimilation (N=620)	Model 4: Dissonant Linguistic Assim. Gap (N=462)
Immigrant vs. Native Youth (dummy: 1=Immigrant Youth)	0.62***	-	-	-
1 st vs. 2 nd /3 rd Immigrant Generation (dummy: 1=1 st Generation)	-	0.89	-	-
Low vs. High Assimilation (dummy: 1=Low-assimilation)	-	-	0.90	-
Small vs. Large Dissonant Assim. Gap (dummy: 1=Small Gap)	-	-	-	1.09
Youth Age (mean scale: 1~5)	1.26***	1.17***	1.19***	1.28***
Youth Gender (dummy: 1=male)	1.69***	1.89***	1.86***	1.73***
Youth Hispanic (dummy: 1=Hispanic)	1.57**	1.69*	1.89**	1.77
Household SES (mean scale: 1~6)	1.03	1.09	1.07	0.96
Concentrated Disadvantage (%)	0.97*	0.96**	0.97*	0.95**
Concentrated Immigration (%)	1.00	1.00	1.00	1.00
Neighborhood Collective Efficacy (mean scale: 1~4)	0.75*	0.64**	0.66*	0.52**
Wald Chi-Square	48.5***	26.3***	26.7***	26.7***
Pseudo R ²	0.0351	0.0331	0.0319	0.0435

Note: All coefficients are presented in the form of odds ratios from logistic regressions with robust standard errors. Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

Table 1-3. Logistic Regression: Outcome of Low Attachment to Family and School

	DV: Low Attachment to Family and School (1=Yes)			
	Model 1: Immigrant Status (N=1034)	Model 2: Generation Status (N=606)	Model 3: Youth Linguistic Assimilation (N=620)	Model 4: Dissonant Linguistic Assim. Gap (N=462)
Immigrant vs. Native Youth (dummy: 1=Immigrant Youth)	0.63**	-	-	-
1 st vs. 2 nd /3 rd Immigrant Generation (dummy: 1=1 st Generation)	-	0.84	-	-
Low vs. High Assimilation (dummy: 1=Low-assimilation)	-	-	0.61**	-
Small vs. Large Dissonant Assim. Gap (dummy: 1=Small Gap)	-	-	-	0.60**
Youth Age (mean scale: 1~5)	1.14**	1.17*	1.11	1.12
Youth Gender (dummy: 1=male)	1.06	1.22	0.99	0.95
Youth Hispanic (dummy: 1=Hispanic)	0.92	1.09	1.52	1.27
Household SES (mean scale: 1~6)	0.92	0.96	0.97	1.08
Concentrated Disadvantage (%)	0.98	0.99	1.00	1.01
Concentrated Immigration (%)	1.01*	1.00	1.01	1.00
Neighborhood Collective Efficacy (mean scale: 1~4)	0.84	0.91	0.86	0.54*
Wald Chi-Square	16.6**	5.4	10.6	11.3
Pseudo R ²	0.0154	0.0098	0.0175	0.0233

Note: All coefficients are presented in the form of odds ratios from logistic regressions with robust standard errors. Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

Table 1-4. Logistic Regression: Outcome of Positive Attitudes toward Delinquency

	DV: Positive Attitudes Toward Delinquency (1=Yes)			
	Model 1: Immigrant Status (N=1036)	Model 2: Generation Status (N=607)	Model 3: Youth Linguistic Assimilation (N=620)	Model 4: Dissonant Linguistic Assim. Gap (N=462)
Immigrant vs. Native Youth (dummy: 1=Immigrant Youth)	0.86	-	-	-
1 st vs. 2 nd /3 rd Immigrant Generation (dummy: 1=1 st Generation)	-	1.01	-	-
Low vs. High Assimilation (dummy: 1=Low-assimilation)	-	-	0.86	-
Small vs. Large Dissonant Assim. Gap (dummy: 1=Small Gap)	-	-	-	1.10
Youth Age (mean scale: 1~5)	1.08	1.14	1.08	1.12
Youth Gender (dummy: 1=male)	1.15	1.00	1.07	1.00
Youth Hispanic (dummy: 1=Hispanic)	1.04	1.04	0.85	0.97
Household SES (mean scale: 1~6)	0.89*	1.03	0.92	0.96
Concentrated Disadvantage (%)	1.03*	1.02*	1.03	1.04
Concentrated Immigration (%)	1.01	1.00	1.00	1.00
Neighborhood Collective Efficacy (mean scale: 1~4)	0.70**	0.74	0.74	0.50***
Wald Chi-Square	42.4***	10.7	10.7	13.6*
Pseudo R ²	0.0332	0.0142	0.0134	0.0245

Note: All coefficients are presented in the form of odds ratios from logistic regressions with robust standard errors. Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

The upper part of Table 1.5 (Model 1) reports the DR estimates of the odds ratios for the linguistic assimilation measure. The linguistic assimilation measure was statistically significantly associated with the attachment outcome. However, the association degree was much smaller when the DR method was used (odds ratio of 0.93) than when the logistic regression method was used (odds ratio of 0.61, which was reported in Table 1-3, Model 3). Holding the other values equal, the odds of having a low level of attachment to family and school were reduced by 7.5% for less-linguistically-assimilated immigrant youth. In other words, DR model indicates that less linguistically-assimilated immigrant youth were more likely to be attached to their family and school.

The lower part of Table 1.5 (Model 2) provides the DR estimates of the odds ratio coefficients when the dissonant linguistic assimilation gap measure was used. All the DR estimates for the three delinquent-risk factor outcomes were statistically insignificant. Interestingly, for the outcome of a low level of attachment to family and school, the DR and logistic regression methods yielded different results in terms of statistical significance. While the odds ratio was statistically significant when the logistic regression was used (as reported in Table 1.3, Mode 4), the odds ratio was no longer statistically significant when the DR method was used.

Table 1-5. Doubly Robust Estimations: Outcome of Three Delinquent Risk Factors

	Treated (dummy =1)	Controlled (dummy =0)	Standard- dized Mean Difference	Standard Deviation Ratios	DR Estimate [Z-value]		
					(1) Delinquency Peers Fraternalization (1=Yes)	(2) Low Attachment to Family and School (1=Yes)	(3) Positive Attitude Toward Delinquency (1=Yes)
Model 1: (N=620) Low vs. High Youth Assimilation (dummy: 1=Low-assimilation)					0.97 [Z= -0.84]	0.93** [Z= -2.19]	0.96 [Z= -0.87]
Youth Age (mean scale: 1~5)	3.26	3.25	0.01	0.99			
Youth Gender (dummy: 1=male)	0.48	0.49	-0.01	1.00			
Youth Hispanic (dummy: 1=Hispanic)	0.95	0.95	-0.03	1.06			
Household SES (mean scale: 1~6)	2.09	2.07	0.03	0.99			
Concentrated Disadvantage (%)	13.5	13.6	-0.03	0.96			
Concentrated Immigration (%)	58.1	57.3	0.06	0.92			
Neighborhood Collective Efficacy (mean scale: 1~4)	2.76	2.74	0.04	<u>0.87</u>			

Model 2: (N=462) Small vs. Large Dissonant Linguistic Assimilation Gap (dummy: 1=Small Dissonant Assimilation Gap)					1.025 [Z= 0.49]	0.941 [Z= -1.42]	1.019 [Z= 0.37]
Youth Age (mean scale: 1~5)	3.19	3.12	0.05	0.93			
Youth Gender (dummy: 1=male)	0.50	0.48	0.04	1.00			
Youth Hispanic (dummy: 1=Hispanic)	0.83	0.78	<u>0.11</u>	0.92			
Household SES (mean scale: 1~6)	2.68	2.70	-0.02	1.00			
Concentrated Disadvantage (%)	12.3	12.6	-0.06	1.05			
Concentrated Immigration (%)	54.3	53.3	0.06	1.00			
Neighborhood Collective Efficacy (mean scale: 1~4)	2.76	2.82	<u>-0.14</u>	1.01			

Note: The underscore indicates that the covariate has more than 10% standardized mean differences or more than 0.1 standard deviation ratios. Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

In summary, when the logistic regression was used, all of the assimilation measures aside from generation status were associated with the greater attachment to family and school. In addition, youth from immigrant households were less likely to fraternize with delinquent friends than their native-born counterparts. There was no association between immigrant youth status and positive attitudes toward delinquency.²²

However, when the DR method was used, the results were less conclusive. The dissonant linguistic assimilation gap measure was no longer statistically significantly associated with the attachment outcome. For the linguistic assimilation measure, the association degree between the measure and the attachment outcome was much smaller in the DR model than in the logistic regression model.

1.5. Robustness Check

1.5.1. Bounding Exercise: Multiple Imputation Analysis of Missing Data

As noted in the previous section, the missing data rates for the assimilation measure calculations were noticeably high. While the rates were relatively low with being 3.7% for immigrant status and 14.8% for generation status, they increased to 24.8% for linguistic assimilation level, and even to 46.1% for family dissonant linguistic assimilation gap. Whether the current statistical results obtained from the complete case observations are reliable depends on the missing data mechanisms (Gelman and Hill, 2007; Little and Rubin, 2002). If the missingness occurred completely at random, the

²² It is also worth noting that, when the statistical significances were ignored, the magnitudes of regression coefficients across all four assimilation measures were consistently small and might be identical.

current statistical inference would not be biased by the missing data. However, if the missingness happened to somewhat or not at random, the inference would be biased.

Investigating missing data structures of the current data revealed that the missingness seemed to occur to somewhat at random, being between "completely at random" and "not at random" (Little and Rubin, 2002). For example, more than one hundred youth who had missing responses to the gender questions also had missing responses to the Hispanic status and linguistic assimilation questions. The missingness in the current data set might have some patterns. If this was true, at least some of the missing data could be imputed from other variable values.

Given the mechanism of missing at random, an additional analysis of missing data imputations provides a type of bounds of the current statistical results. In this paper, the multivariate normal model (Little and Rubin, 2002) was used for multiple imputations of missing data. In this model, it was assumed that all the variables having any missing value could be simultaneously predicted by the complete case variables having no missing value through a link of multivariate normal distribution.²³ Using the Markov Chain Monte Carlo (MCMC) simulation technique, the missing observations were completely imputed by randomly chosen values from the fitted multivariate normal distribution. After the imputation, the all variables had the numbers of total sample

²³ For the current missing data analysis, the missing data variables included all the four assimilation measures, the outcomes of low attitudes to family and positive attitudes toward delinquency, youth age, gender, Hispanic, household SES, and neighborhood collective efficacy, while the complete case variables included the other outcomes of delinquent peer fraternization and victimization, and 12 neighborhood structural variables (%black, %population aged 10~19, %female-headed family, %household that received food-stamp assistance in the past 12 months, %unemployed, %household whose income is below the poverty line, %Hispanic, %foreign-born, %moved within the past 1 year, %those who speak languages other than English at home, %college graduates, %household that receive social security).

observations, being 1,435.²⁴ This imputation process was iterated 500 times. Each of the iterations produced one completely imputed hypothetical data set, thus now creating 500 data sets.

Table 1.6 reports the logistic regression odds ratio distributions for the four assimilation measures, using the 500 hypothetical data sets obtained from the multiple imputation process. Among the newly obtained 500 odds ratios, the minimum and maximum provided possible bounds of the current logistic regression odds ratios.

For the assimilation measure of immigrant status, all the reported bounds of odds ratio were less than 1, implying that if the missing imputation model was correct, even when there had been no missing value for all observations, immigrant status would be likely to be associated with a low level of delinquent risk factor development. It is notable that the original odd ratios for the outcomes of delinquent peer fraternization and low attachment to family and school were close to the extreme values of the bounds, suggesting a possibility that the logistic regression results might be biased.²⁵

The ranges of the possible odds ratios were from less than 1 to greater than 1 for all the four outcomes across the two language-related measures. For the outcome of low attachment to family and school, the original odds ratios at the complete case analysis were 0.61 for the youth linguistic assimilation measure, and 0.60 for the family dissonant linguistic assimilation gap measure (as in Table 1.3). Now the possible bound range

²⁴ One observation that did not have location information was dropped from the 907 total objects.

²⁵ Interestingly, the original odds ratios from the complete case analyses were excluded from the ranges by a slight margin. This could happen because the missing data analysis in this section relied on the simulation process, called the MCMC, which generated random numbers for creating the hypothetical data sets.

suggested by the missing data analysis was between 0.61 and 1.19 for the former, and between 0.48 and 1.05 for the latter. Noticeably, the original 0.61 odds ratio for the youth linguistic assimilation measure was close to the extreme value of the bounds.

Table 1-6. Logistic Regression Odds Ratios Using 500 Hypothetical Missing-Imputed Data Sets

Outcome	Original OR obtained without Imputation	Bounds from All the 500 New Odds Ratio Estimates		
		Mean (SD)	(Minimum,	Maximum)
Model 1. for the Measure of Immigrant Status (N=1,434)				
(1) Delinquent Peer Fraternization	0.62**	0.70 (0.03)	(0.64,	0.81)
(2) Low Attachment to Family and School	0.63**	0.76 (0.03)	(0.64,	0.85)
(3) Positive Attitudes toward Delinquency	0.86	0.88 (0.04)	(0.71,	0.99)
Model 2. for the Measure of Generation Status (N=796)				
(1) Delinquent Peer Fraternization	0.89	0.86 (0.11)	(0.61,	1.30)
(2) Low Attachment to Family and School	0.84	0.91 (0.13)	(0.64,	1.33)
(3) Positive Attitudes toward Delinquency	1.01	1.03 (0.12)	(0.73,	1.45)
Model 3. for the Measure of Low Youth Linguistic Assimilation Level (N=906)				
(1) Delinquent Peer Fraternization	0.90	0.89 (0.11)	(0.54,	1.24)
(2) Low Attachment to Family and School	0.61**	0.82 (0.09)	(0.61,	1.19)
(3) Positive Attitudes toward Delinquency	0.86	0.93 (0.09)	(0.67,	1.21)
Model 4. for the Measure of Small Family Dissonant Linguistic Assimilation Gap (N=906)				
(1) Delinquent Peer Fraternization	1.09	1.15 (0.17)	(0.73,	1.79)
(2) Low Attachment to Family and School	0.60**	0.75 (0.10)	(0.48,	1.05)
(3) Positive Attitudes toward Delinquency	1.10	1.06 (0.13)	(0.73,	1.55)

Note: Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

1.5.2. Revising Generation and Linguistic Assimilation Measures with Those Native-Born

Due to the large number of missingness, the sample sizes for the logistic regression and the DR estimation analyses were decreased. If there had been no missingness, the sample size for the analysis using the generation status measure would be 797, and that using the linguistic assimilation measure would be 907. However, in the current paper, the sample sizes were reduced to 679 and 682, respectively, due to the missingness. The decreases in sample sizes led to a loss of the statistical power to detect the associations between assimilation measures and delinquent risk factor outcomes.

Utilizing the fact that immigrant youth who have a high level of assimilation are close to native-born youth, the sample sizes for the analyses using the generation status and linguistic assimilation measures could be increased.²⁶ The generation status and linguistic assimilation measures were revised in the following way. With the 585 native-born youth added to the control group who were 2nd or 3rd generation immigrant youth, the revised generation status measure had 135 1st immigrant generation youth (treatment group) and 1,129 2nd/3rd immigrant generation or native-born youth (control group). Also, 512 native-born youth who did not answer the linguistic assimilation questions in the survey were added to the revised linguistic assimilation measure, thus having 382 immigrant youth who had a low level of linguistic assimilation (treatment group) and 812

²⁶ However, the case using the dissonant linguistic assimilation gap measure was excluded from this section. Those who have small linguistic assimilation gaps with their parents are less-assimilated immigrant youth and native-born youth, while those who have large linguistic assimilation gaps with their parents are more-assimilated immigrant youth. However, because the less-assimilated immigrant youth tend to be significantly different from native-born youth, they cannot be combined together.

native-born or immigrant youth who had a high/perfect level of linguistic assimilation (control group).

Table 1-7 reports the logistic regression results for the two assimilation measures and the three delinquent risk factor outcomes. The results were almost the same as those in the previous tables when the native-born youth were not added. Among the six outcomes, only the youth's own linguistic assimilation measure was associated with low level of attachment to family and school (Column (4)). However, the DR estimation results for the youth linguistic assimilation measure in Table 1-8 were different from those of the logistic regressions. The linguistic assimilation measure from the DR estimation shows that a low level of linguistic assimilation is associated with a low level of delinquent peer fraternization and not with differences in attachment to family and school.

Table 1-7. Robustness Check: Logistic Regression Outcomes with Revised Assimilation Measures

	DV: Delinquent Peer Association (1=Yes)		DV: Low Attachment to Family and School (1=Yes)		DV: Positive Attitudes Toward Delinquency (1=Yes)	
	(1) Generation Status (N=1,037)	(2) Youth Linguistic Assimilation (N=982)	(3) Generation Status (N=1,035)	(4) Youth Linguistic Assimilation (N=981)	(5) Generation Status (N=1,035)	(6) Youth Linguistic Assimilation (N=981)
1 st vs. (2 nd /3 rd + Native-born) Generation (dummy: 1=1 st Generation)	0.79	-	0.74	-	0.95	-
Low vs. (High + Native-born) Assimilation (dummy: 1=Low-assimilation)	-	0.78	-	0.58***	-	0.80
Youth Age (mean scale: 1~5)	1.25***	1.26***	1.14*	1.13**	1.08	1.09
Youth Gender (dummy: 1=male)	1.71***	1.66***	1.07	1.01	1.14	1.19
Youth Hispanic (dummy: 1=Hispanic)	1.31	1.48*	0.76	0.85	0.97	1.07
Household SES (mean scale: 1~6)	1.06	1.05	0.95	0.93	0.90	0.87
Concentrated Disadvantage (%)	0.97*	0.97	0.98	0.98	1.03*	1.03
Concentrated Immigration (%)	1.00	1.00	1.01*	1.01*	1.01	1.00
Neighborhood Collective Efficacy (mean scale: 1~4)	0.79	0.84	0.86	0.88	0.71**	0.70**
Wald Chi-Square	41.22***	40.14***	13.93*	17.45**	42.18***	44.23***
Pseudo R ²	0.0301	0.0304	0.0121	0.0172	0.0326	0.0360

Note: All coefficients are presented in the form of odds ratios from logistic regressions with robust standard errors. Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

Table 1-8. Robustness Check: DR Estimation Outcomes with Revised Assimilation Measures

	Treated (dummy =1)	Controlled (dummy =0)	Standard- dized Mean Difference	Standard Deviation Ratios	DR Estimate [Z-value]		
					(1) Delinquency Peers Fraternalization (1=Yes)	(2) Low Attachment to Family and School (1=Yes)	(3) Positive Attitude Toward Delinquency (1=Yes)
Low vs. High Youth Assimilation (dummy: 1=Low-assimilation)					0.89** [Z= -2.20]	0.93 [Z= -1.25]	0.99 [Z= -0.27]
Youth Age (mean scale: 1~5)	3.26	3.13	<u>0.11</u>	0.95			
Youth Gender (dummy: 1=male)	0.48	0.49	-0.01	1.00			
Youth Hispanic (dummy: 1=Hispanic)	0.95	0.94	0.03	0.95			
Household SES (mean scale: 1~6)	2.09	2.05	0.05	1.07			
Concentrated Disadvantage (%)	13.47	13.23	0.05	1.02			
Concentrated Immigration (%)	58.12	57.34	0.06	0.93			
Neighborhood Collective Efficacy (mean scale: 1~4)	2.76	2.77	0.03	<u>0.88</u>			

1.6. Discussion and Conclusion

The current paper investigated the associations between assimilation measures and delinquent risk factors for immigrant youth. Compared to the extant studies, the current paper demonstrated the following differences. First, this study focused on three delinquent-risk factors rather than measures of delinquent behavior as the outcome variables. Second, all the four different assimilation measures—including immigrant status, linguistic assimilation, generation status, and the linguistic assimilation gap between youth and parents—were used as the explanatory variables. Third, this paper compared estimates from logistic regressions and doubly robust (DR) methods to explore the associations.

The current study does not address any causal relationship between assimilation and delinquent risk factors. The logistic regression design is naturally an exploratory design. Although the propensity-score modeling in the DR estimation method can help address problems of endogeneity issues such as selection bias, it cannot fully solve the problems because the propensity-score modeling cannot control for unobserved variables. Therefore, the result should be interpreted only as associations between variables. Even when an assimilation level is associated with a delinquent risk factor, it does not necessarily mean that the assimilation level causes the delinquent risk factor development.

In summary, the logistic regression and DR estimation results are partially supportive of the positive association between assimilation levels and likelihoods of delinquent-risk development. Across all the assimilation measures and all the estimation methods, a low level of assimilation tended to be associated with a high level of

attachment to family and school for immigrant youth. However, the associations between assimilation levels and the delinquent peer fraternization outcome remained inconclusive because of the inconsistency in statistical significances. The delinquent-risk factor of positive attitudes toward delinquency had no statistically significant association with any assimilation measure.

Among the four assimilation measures, the immigrant status seems to be a relatively strong predictor of the delinquent risk factor outcomes. In general, immigrant youth were more likely to avoid delinquent peers and to feel more attached to their family and school than their native-born counterparts. Meanwhile, the generation-status measure had no statistically significant association with any delinquent risk factor development. The two language-related assimilation measures had mixed results. The results for these assimilation measures had inconsistencies in statistical significances and magnitudes of coefficients. In particular, with the linguistic assimilation measure and the attachment outcome, the estimated odds ratio magnitudes shrunk dramatically in moving from the logistic regression to the DR estimation.

Several implications arise from the statistical results. One is that the four single-dimension assimilation measures, especially the language-related ones, may be relatively weak predictors of delinquent-risk factor development. The odds ratio magnitudes of these four measures were relatively small, ranging from 0.5 to 2.0, with the majority being just around 1.0. In particular, the linguistic assimilation measure seems to be less effective in predicting delinquent risk factors outcomes than most extant studies argue. It meshes with Morenoff and Astor's (2006) report that their linguistic assimilation measure

was not as strong a predictor of crime outcomes as other measures, such as the age when immigrants arrived in the U.S.

One possible explanation for this weakness of the current assimilation measures is that their construct validities are low due to their single dimensions. Assimilation is a total transition process of an immigrant's internalized cognitive and behavioral systems, including thoughts, beliefs, and attitudes (Finch et al., 2000; Vega and Gil, 1998). Those internal cognitive and behavioral changes for immigrant youth may be too complicated to be captured by quantitative measures (MacDonald and Saunders, 2012). In that sense, composite measures that gauge multiple dimensions of assimilation may be more useful in predicting delinquent risk factors than single-dimension measures.

However, it is worth noting that the current study results are consistent with the "Hispanic Paradox" or "Latino Paradox" (Palloni and Arias, 2004; Sampson, 2006 and 2008; Scribner, 1996). And while the study is an exploratory design, the results do not lend any support to the idea that immigrant youth are associated with more negative outcomes. Rather, the results suggest, if any, that less-assimilated immigrant youth tend to be associated with less delinquent risk.

Another implication of this study is that positive attitudes toward delinquency had no statistically significant association with any of the assimilation measures, which is against what the segmented assimilation theory expects. Assuming that subcultural deviant norms can be represented by the positive attitudes toward delinquency in the current study, the results indicate that even highly assimilated youth do not necessarily

absorb the subcultural deviant norms (see Alba and Nee, 1997; Perlmann and Waldinger, 1997 for the similar argument).

There exist a number of limitations in the current paper. One is still the missingness, related to the willingness to report. Although the missing data analysis in Section 5 was based on the strong assumption that the missingness in the current data set occurred "at random," thus being predictable from other variables, there is still the possibility that the missingness might have occurred "not at random," depending solely on the unobserved own nature of the predictors and being unpredictable from other variables (Gelman and Hill, 2007). For example, it might be that immigrant youth who were less assimilated had a low level of willingness to report their low level of assimilation. If this is the case, the current statistical inference only with the complete sample cases might be biased, being different from what could have been obtained when there was no missing value.

Also, because the majority of immigrant samples were Hispanic and most of them lived in specific inner city enclaves in Los Angeles, the generalizability of the current study results is limited (Bersani, Loughran, and Piquero, 2013). Hispanic immigrants who newly migrated to their ethnic enclaves such as Los Angeles could expect some help from their ethnic neighbors, and the strong ethnic connection to the existing community would provide some protections for new immigrants, such as informal social controls or an economic opportunity. However, the results may not hold for immigrants who are following the recent trend of settling into new rural destinations where ethnic enclaves do not currently exist (Johnson and Lichter, 2008).

In addition, the data used for the current paper were cross-sectional, leaving most unobservable confounding factors uncontrolled in the regression modeling. Using self-reported data is another source of potential bias in the results, such as exaggeration and intentionally not answering particular questions.

Nevertheless, this paper has some noticeable strength. It empirically addressed the association between multiple assimilation measures, delinquent-risk factors, with a rare set of data that have language-based assimilation measures for both parents and youth from immigrant households. Also, this is the first study that has attempted to address associations between assimilation and delinquent-risk factors using a propensity core method called the DR method.

APPENDIX 1-A. Survey Question Item Statements

1. Dependent Variables

(1) Delinquent Risk Factor: Delinquent Peer Fraternization

: "In the past 12 months, that is since July 2007, have ANY of your friends or people you regularly socialize with or hang out with ..."

- a. Purposely damaged or destroyed property that did not belong to them?
- b. Taken money or property that didn't belong to them?
- c. Broken into a house or building to steal something or just to look around?
- d. Taken a car that didn't belong to them?
- e. Hit someone or gotten into a physical fight?
- f. Hurt someone badly enough that they needed bandages or a doctor?
- g. Used a knife or gun or some other weapon like a club to get something from a person?
- h. Drunk alcohol at least once a week?
- i. Used marijuana at least once a month?

(2) Delinquent Risk Factor: Low Attachment to Family and School

: "How much do the following statements describe how you feel about your family/school?"

A. Family

- a. Your parents care about you.
- b. People in your family understand you.
- c. You and your family have fun together
- d. Your family pays attention to you.

B. School

- a. You are close to people at school.
- b. You are part of your school.
- c. You are happy to be at school.
- d. The teachers at school treat you fairly.

(3) Delinquent Risk Factor: Positive Attitude toward Delinquency

: "Please tell me how much you agree with the following statements about your neighborhood...."

- a. If someone disrespects another youth, he or she has to fight to teach that person not to disrespect them. Do you....
- b. It is important that a youth act tough; otherwise he or she will become a victim. Do you....
- c. Sometimes youth have to threaten people in order to get treated fairly. Do you....
- d. Youth give thugs or gang members a lot of respect. Do you....
- e. Youth don't respect other youth who are afraid to fight. Do you....
- f. Youth don't think much about their future. Do you....

2. Independent Variables

(1) English Assimilation

- a. In what language are the movies, TV, and radio programs you want to watch or listen to?
- b. What language(s) do you speak?
- c. What language(s) do you read?
- d. With whom (speaks ...) do you socialize?

3. Control Variables

(1) Collective Efficacy

A. Social Cohesion

: "Now I am going to read some statements that may or may not be true of your neighborhood. For each statement, please tell me whether you strongly disagree, disagree, agree, or strongly agree...."

- a. People around here are willing to help their neighbors.
- b. This is a close-knit neighborhood.
- c. People in this neighborhood can be trusted.

- d. People in this neighborhood generally don't get along with each other.
- e. People in this neighborhood do not share the same values
- f. Parents in this neighborhood know their children's friends.
- g. Adults in this neighborhood know who the local children are.
- h. Parents in this neighborhood generally know each other.
- i. People in this neighborhood are willing to do favors for each other like watching each other's children, helping with [is something missing here?]

B. Informal Social Control

: "For the next few questions, please tell me whether it is very unlikely, unlikely, likely, or very likely that your neighbors would do something if..".

- a. Children were skipping school and hanging out on a street corner.
- b. If children were spray-painting graffiti on a sidewalk or building.
- c. If children were showing disrespect to an adult.
- d. If a fight broke out in public.
- e. If a youth gang was hanging out on the street corner selling drugs and intimidating people.
- f. If a local school near your home was threatened with closure due to budget cuts.

APPENDIX 1-B. Corresponding OLS Regression Results

Table 1-9. OLS Regression: Outcome of Delinquent Peer Fraternization

	DV: Delinquent Peer Fraternization Index (1~9)			
	Model 1: Immigrant Status (N=1036)	Model 2: Generation Status (N=607)	Model 3: Youth Linguistic Assimilation (N=620)	Model 4: Dissonant Linguistic Assim. Gap (N=462)
Immigrant vs. Native Youth (dummy: 1=Immigrant Youth)	-0.54*** (0.16)	-	-	-
1 st vs. 2 nd /3 rd Immigrant Generation (dummy: 1=1 st Generation)	-	-0.21 (0.15)	-	-
Linguistic Assimilation Index (mean scale: 1 ~ 5)	-	-	0.22 (0.14)	-
Dissonant Assimilation Gap Index (mean scale: -4 ~ +4)	-	-	-	0.04 (0.09)
Youth Age (mean scale: 1~5)	0.13** (0.04)	0.07 (0.05)	0.08* (0.05)	0.13** (0.05)
Youth Gender (dummy: 1=male)	0.27** (0.11)	0.25* (0.14)	0.26** (0.14)	0.23 (0.15)
Youth Hispanic (dummy: 1=Hispanic)	0.17 (0.16)	0.11 (0.25)	0.31 (0.23)	0.26 (0.25)
Household SES (mean scale: 1~6)	-0.05 (0.05)	0.04 (0.07)	0.04 (0.07)	0.03 (0.08)
Concentrated Disadvantage (%)	-0.02 (0.01)	-0.007 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Concentrated Immigration (%)	0.001 (0.004)	0.003 (0.005)	0.004 (0.005)	0.003 (0.005)
Neighborhood Collective Efficacy (mean scale: 1~4)	-0.35** (0.14)	-0.19 (0.19)	-0.19 (0.19)	-0.55*** (0.20)
F-Statistic	3.75***	1.08	1.75*	2.52**
R ²	0.0341	0.0124	0.0169	0.0323

Note: Robust standard errors are in parenthesis. Statistical Significance Level: * <0.1, ** <0.05, *** <0.01

Table 1-10. OLS Regression: Outcome of High Attachment to Family and School

	DV: High Attachment to Family and School Index (1~5)			
	Model 1: Immigrant Status (N=1036)	Model 2: Generation Status (N=607)	Model 3: Youth Linguistic Assimilation (N=620)	Model 4: Dissonant Linguistic Assim. Gap (N=462)
Immigrant vs. Native Youth (dummy: 1=Immigrant Youth)	0.10** (0.04)	-	-	-
1 st vs. 2 nd /3 rd Immigrant Generation (dummy: 1=1 st Generation)	-	0.04 (0.05)	-	-
Linguistic Assimilation Index (mean scale: 1 ~ 5)	-	-	-0.15*** (0.05)	-
Dissonant Assimilation Gap Index (mean scale: -4 ~ +4)	-	-	-	-0.07** (0.03)
Youth Age (mean scale: 1~5)	-0.04*** (0.01)	-0.05*** (0.02)	-0.04** (0.02)	-0.04** (0.02)
Youth Gender (dummy: 1=male)	-0.07** (0.03)	-0.11** (0.05)	-0.06 (0.04)	-0.06 (0.05)
Youth Hispanic (dummy: 1=Hispanic)	0.02 (0.05)	-0.02 (0.08)	-0.03 (0.08)	-0.01 (0.09)
Household SES (mean scale: 1~6)	0.02 (0.02)	0.02 (0.03)	0.009 (0.02)	-0.01 (0.03)
Concentrated Disadvantage (%)	0.002 (0.004)	-0.001 (0.005)	-0.003 (0.005)	-0.004 (0.006)
Concentrated Immigration (%)	-0.002* (0.001)	-0.001 (0.002)	-0.002 (0.002)	-0.001 (0.002)
Neighborhood Collective Efficacy (mean scale: 1~4)	0.04 (0.04)	0.05 (0.06)	0.06 (0.05)	0.11* (0.06)
F-Statistic	3.37***	2.17**	2.81***	1.88*
R ²	0.0266	0.0322	0.0368	0.0356

Note: Robust standard errors are in parenthesis. Statistical Significance Level: * <0.1, ** <0.05,

*** <0.01

Table 1-11. OLS Regression: Outcome of Negative Attitudes toward Delinquency

	DV: Negative Attitudes toward Delinquency Index (1~4)			
	Model 1: Immigrant Status (N=1036)	Model 2: Generation Status (N=607)	Model 3: Youth Linguistic Assimilation (N=620)	Model 4: Dissonant Linguistic Assim. Gap (N=462)
Immigrant vs. Native Youth (dummy: 1=Immigrant Youth)	0.08 (0.05)	-	-	-
1 st vs. 2 nd /3 rd Immigrant Generation (dummy: 1=1 st Generation)	-	0.06 (0.06)	-	-
Linguistic Assimilation Index (mean scale: 1 ~ 5)	-	-	-0.10* (0.05)	-
Dissonant Assimilation Gap Index (mean scale: -4 ~ +4)	-	-	-	-0.04 (0.03)
Youth Age (mean scale: 1~5)	-0.01 (0.01)	-0.02 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Youth Gender (dummy: 1=male)	-0.05 (0.04)	-0.01 (0.05)	-0.02 (0.05)	-0.001 (0.05)
Youth Hispanic (dummy: 1=Hispanic)	-0.08 (0.05)	-0.06 (0.08)	-0.01 (0.08)	-0.02 (0.08)
Household SES (mean scale: 1~6)	0.03 (0.02)	-0.01 (0.02)	0.02 (0.02)	-0.01 (0.03)
Concentrated Disadvantage (%)	-0.01*** (0.005)	-0.01** (0.006)	-0.01** (0.006)	-0.01* (0.007)
Concentrated Immigration (%)	-0.001 (0.001)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Neighborhood Collective Efficacy (mean scale: 1~4)	0.13*** (0.05)	0.13** (0.07)	0.13** (0.06)	0.27*** (0.07)
F-Statistic	11.00***	2.50**	2.42**	3.27***
R ²	0.0792	0.0348	0.0334	0.0587

Note: Robust standard errors are in parenthesis. Statistical Significance Level: * <0.1, ** <0.05,

*** <0.01

This appendix section reports OLS regression results with continuous delinquent risk factor outcomes and continuous language-related assimilation measures. These OLS regression results are almost the same as those from the logistic regressions.

Chapter 2 . THE EFFECT OF THE INCREASED ALCOHOL AVAILABILITY ON CRIME: A QUASI-EXPERIMENTAL ANALYSIS OF THE REPEAL OF THE SUNDAY OFF-PREMISE LIQUOR SALES BAN

Abstract

This paper investigates the effect of increased alcohol availability on localized crime rates. With the repeal of a long-standing "Blue Law" in 2003, Pennsylvania allowed a portion of the state-run liquor stores in Philadelphia to remain open on Sunday, while keeping the remainder of the stores in the city closed. This creates a treatment area group in which alcohol availability increased and a comparison group in which Sunday liquor sales were still prohibited. Localized crime rates for each store were gauged using counts of crime incidents occurring within multiple radii of variable sizes around each of the liquor stores. To capture the effect of the policy change, a difference-in-difference-in-difference (prior vs. post-repeal; Sunday vs. non-Sundays; treatment vs. control group) method was employed. This study consistently found some evidence for local geographical crime pattern changes after the repeal. While total crime incidents occurring in the immediate vicinity of the Sunday-open state liquor stores significantly increased, there was a slight decrease in total crime incidents occurring in the areas farther away from the stores, which suggests an existence of a local crime displacement, or "attraction," effect. In addition, the study results showed that the crime attraction effect was present only in low socioeconomic status (SES) areas.

2.1. Introduction

There was a long tradition of prohibiting commercial transactions on Sundays, known as "Blue Laws".²⁷ Many of them were eliminated by the late twentieth century (Lovenheim and Steefel, 2011), but alcohol restrictions remain one of the few areas where "Blue Laws" are still in effect. However, in the mid-1990s when states experienced serious budget deficits, the "Blue Law" states started to reconsider lifting Sunday alcohol sales bans. While repealing the bans is to raise excise tax revenues, public health and safety concerns are often cited as the basis for continuing their use (Carpenter and Eisenberg, 2009; Heaton, 2012; McMillan and Lapham, 2006; Lovenheim and Steefel, 2011; Stehr, 2010).²⁸ Currently 38 states permit varying levels of Sunday off-premise liquor sales and benefit from the resulting increased tax revenues.^{29 30}

Scientific studies also join the debate. As the first to implement the repeal in 1995, New Mexico's initiative received the most scrutiny. However, the New Mexico studies with alcohol-related traffic fatality and crash outcomes were conflicting and inconclusive (Maloney and Rudbeck, 2006 and 2009; McMillan and Lapham, 2006), and so were evaluations of the other states' repeals (Lovenheim and Steefel, 2011; Stehr, 2010). In

²⁷ See Lovenheim and Steefel (2011) for a historical summary of the general "Blue Laws" in the U.S.

²⁸ According to the Distilled Spirits Council of the United States (hereafter DISCUS), 17 states have repealed their blue laws since 1995. These states are Arkansas (2009), Colorado (2008), Connecticut (2012), Delaware (2003), Georgia (2011), Idaho (2004), Kansas (2004), Kentucky (2004), Massachusetts (2003), New Mexico (1995), New York (2003), Ohio (2004), Oregon (2002), Pennsylvania (2003), Rhode Island (2004), Virginia (2004), and Washington (2005). Some rolled out their repeals in multiple steps.

²⁹ In other words, as of June 2013, twelve states and the District of Columbia continue to prohibit the off-premise sale of alcohol on Sunday. These states are Alabama, Indiana, Minnesota, Mississippi, Montana, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Utah, and West Virginia.

³⁰ Eighteen states permit Sunday off-premise liquor sales statewide, 16 states entrust permission to local authorities, and 4 states, including Pennsylvania, control liquor sales based on certain geographic or demographic characteristics (<http://www.discus.org/policy/sunday/>, accessed on June 5, 2013).

general, beyond the traffic-related outcomes, comprehensive effects of the "Blue Law" repeals on public health and safety were still required.

In response to this demand, Heaton (2012) was the first that investigated the repeal effects on general crime rates. Utilizing Virginia's rolling-out Sunday off-premise liquor sale permissions in selected city/county jurisdictions since 2004, he showed that alcohol-related serious crimes and misdemeanor offenses increased in the jurisdictions permitting off-premise liquor sales on Sundays through a difference-in-difference-in-difference design. In addition, no geographical displacements of crime incidents were detected; jurisdictions surrounding the study areas experienced no significant crime increase after the repeal on Sundays. It was concluded that the Virginia's "Blue Law" repeal had at least some effects on certain types of crime on the Sundays in which liquor was sold in the ban-lifted jurisdictions.

However, the jurisdiction-level analysis may not fully describe the dynamic effects of the repeal on crime. Given large sizes of jurisdictions, local effects of the repeal may be concealed behind the jurisdiction-level aggregation (Barr and Pease, 1990; Clarke and Mayhew, 1988; Cornish and Clarke, 1987; Teh, 2008). For example, crime may move locally toward alcohol outlets from other places within a jurisdiction, as the outlets work as crime "generators" or "attractors" (Brantingham and Brantingham, 1993 and 1995). These local dynamic effects, if any, will be best captured by measures of smaller geographical units of analysis than in large jurisdictions (Gruenewald, Millar, and Roeper, 1996; Scribner et al., 1999).

The current study uses different-sized surrounding areas of liquor stores as its units of analysis. Pennsylvania repealed its "Blue Laws" since 2003, allowing only a part of its state-run stores to sell liquor off-premise on Sundays. Utilizing this statewide change as a quasi-experiment leverage, this study traces local variations of crime incidents occurring around ban-repealing liquor stores in Philadelphia. The study results show that areas in the immediate vicinity of liquor stores witnessed an increase in total crime incidents after the repeal, and that part of the increase might be displaced from areas farther away from the stores. The results indicate a local crime-redistribution effect of the repeal within a jurisdiction, with the off-premise liquor stores functioning as crime attractors.

The remainder of this paper is organized as follows. The following second section provides background information on Pennsylvania's repeal of the ban on Sunday off-premise liquor sales, reviews of extant studies that examine the association between the alcohol availability increase and crime, and some possible mechanism explanations. Note that the current study tests the effects of alcohol availability, including a congregation effect of motivated offenders and potential victims to the availability-increased areas, not necessarily the effects of alcohol use. The next section describes aspects of the study design, including data sources, variables, and double-difference (hereafter DD) and triple-difference (hereafter DDD) models. The fourth section provides the primary DD and DDD results, along with the crime local displacement or attraction effects of the repeal. Whether the repeal effects vary depending on the neighborhoods' socioeconomic status (SES) is also briefly investigated. The fifth section examines the robustness of the

results. Finally, a summary, implications, and limitations of the current study are discussed.

2.2. Background and Literature Review

2.2.1. Pennsylvania "Blue Law" Repeal

The Pennsylvania Liquor Control Board (hereafter PLCB) has the authority to issue alcohol-sale licenses, supervise alcohol-sale licensees, and directly run state-monopolistic liquor stores called Wine & Spirits stores (hereafter W&S) in Pennsylvania. Its authority is ascribed by the Pennsylvania Liquor Code.³¹ Liquor is defined as "*mean and include any alcoholic, spirituous, vinous, fermented or other alcoholic beverage, or combination ... which contain more than one-half of one per cent of alcohol by volume, except pure ethyl alcohol and malt or brewed beverages,*" while malt or brewed beverages are "*any beer, lager beer, ale, porter or similar fermented malt beverage containing one-half of one per cent centum or more of alcohol by volume*" (Liquor Code §1-102).

The Liquor Code has been modified several times since its initial implementation. One notable revision was enacted on December 9, 2002, as Act 212 of 2002, repealed Pennsylvania's "Blue Law" that affected liquor sales. Liquor Code §3-304 was amended to allow the PLCB to open up to 10% of its W&S stores from 12:00 PM to 5:00 PM on Sundays. Prior to this change, all W&S stores were closed on Sundays. When the Act

³¹ The Liquor Code complies with Pennsylvania Act 21 of April 12, 1951, P.L. 90, which was reenacted by Act 14 of June 29, 1987, P.L. 32, and the PLCB Regulations and the Administrative Code of 1929, Sec. 614A.

took effect 60 days later, on February 9, 2003, the 10% of W&S stores included eight located in Philadelphia.

On December 8, 2004, the Liquor Code was amended further by Act 239 of 2004 to allow 25% of the W&S stores in Pennsylvania to open on Sundays; this became effective on February 7, 2005. Since then, the total number of Sunday-open W&S stores located in Philadelphia County has varied from year to year: with a total of 77 different stores open at some point on a Sunday. For the period of observations in this study we have no W&S stores out of 65 open January 1, 1998, 8 out of 65 open as of January 1, 2004 (12.3%), 15 out of 62 open as of January 1, 2006 (24.2%), and 19 out of 54 open as of January 1, 2011 (25.8%). Detailed information on the W&S stores located in Philadelphia is provided in Appendix 2-A.

It is worth emphasizing that the Liquor Code revisions in 2003 and 2005 repealed only the off-premise liquor sales ban of the W&S stores. These modifications did not affect on-premise liquor sales or Sunday on- and off-premise malt and brewed beverage sales by hotels and restaurants (and clubs only for on-premise sales) that hold liquor licenses. Even before 2003, these licensees could sell liquor and malt or brewed beverages on Sundays by purchasing a special permit for on-premise consumption (Liquor Code §4-406), and also could sell malt or brewed beverages on Sundays for off-premise consumption (Liquor Code §4-407). Therefore, the current paper dealing with the Pennsylvania repeal limits its interest to Sunday "off-premise" "liquor" sales by state-run W&S stores.

2.2.2. Literature Review

Alcohol is clearly correlated with crime, but the nature of the causal relationship remains ambiguous. Increased availability of alcohol boosts consumption, which, in turn, raises crime rates (Carpenter and Dobkin, 2011). However, possible endogeneity due to unobserved factors prevents us from directly addressing the causal association between alcohol and crime. For example, both alcohol use and crime may be manifestations of a similar underlying social or psychological process, such as risk-preferred temperament.

Therefore, it is important to identify empirically whether and which crime-prevention policies that control alcohol availability can lead successfully to a reduction in crime rates. A variety of types of alcohol control policies have been assessed for their correlation with crime. In addition to the traditional empirical studies of excise-tax policy and the minimum-age restriction, location-and time-restriction approaches have also been examined. According to Carpenter and Dobkin (2011), the empirical evidence is strongest on the effects of age restrictions on crime. In contrast, there is some limited evidence that alcohol excise taxes reduce crime, and only limited evidence that location and time restrictions on alcohol availability affect crime.

The empirical studies on the effects of the traditional excise-tax policy and the minimum-age restriction on crime are not covered in detail in this paper's review. Carpenter and Dobkin (2011) provide a comprehensive review of studies on age restrictions and crime. Among the extant studies on the effect of the excise-tax policy on crime, one study recently published deserves attention. Instead of directly estimating the tax increase policy effects on crime outcomes, Cook and Durrance (2013) use per capita

alcohol consumption as a moderator of the 1991 nationwide federal tax increase effect on crime outcomes. This method provides heterogeneity in policy effects for states, while taking advantage of the uniform national tax increase on alcohol in 1991 as an exogeneity shock. They found that the excise-tax increase reduced crime and fatal crash outcomes, and the benefit of the reduction is larger for "wet" states having a generous alcohol policy than for "dry" states having a strict alcohol policy.

The weak evidence for the relationship between location and time restrictions of alcohol availability and crime can be explained in part by the conflict between two theoretical perspectives on the direction of any effect (Heaton, 2012). One perspective would argue there should be little, if any, effect of time and location restrictions. People can buy and consume alcohol at times other than the restricted ones (known as "inter-temporal substitutability"). Individuals who live close to areas that freely allow alcohol can travel to these locations to purchase alcoholic beverages (referred to as "geographical substitutability"). Also, if a certain type of alcohol is highly regulated, people can buy and consume other types of alcohol that are not as tightly controlled (known as "product substitutability"). Thus, an alcohol policy with a single type of restriction on location or time is unlikely to reduce an overall volume of alcohol consumption and, thus, crime rates.

In contrast, situational crime-prevention theorists suggest that location and time restrictions should be effective in crime reduction (Cornish and Clarke, 1987; Eck, 1993). According to them, any intervention may cause potential offenders and victims to change their behaviors, and the behavioral changes, in turn, may lead to crime displacement,

impacting when, where, how, and which crimes occur (Barr and Pease, 1990; Hakim and Rengert, 1981; Reppetto, 1976). However, this situational process of crime does not necessarily result in crime displacement when potential offenders consider benefits and costs of crime perpetration. When a change in the crime environment occurs in such a way that the costs exceed benefits, potential offenders do not commit crime that they would commit if it were not for the cost and benefit structure change of crime perpetration. Therefore, even when alcohol consumption is displaced by any intervention restricting time and location of alcohol availability, an overall volume of crime may be reduced since crime is less likely to be displaced due to potential offenders' cost-benefit calculations.

The empirical results of alcohol-availability restrictions show support for both alcohol substitution and situational contingency of crime. First, empirical studies on the relationship between crime and locational alcohol availability that are commonly measured by the alcohol outlet density show mixed results. For example, Scribner and his colleagues show that one more alcohol outlet addition was associated with an increase in violent assaults in Los Angeles (Scribner, Mackinnon, and Dwyer, 1995) and an increase in homicide rates in New Orleans (Scribner et al., 1999). Gruenewald et al. (2006) report that assaults were more common at off-premises than at on-premise establishments and that a positive association between bar density and assault rates was consistent only in low-income, poor, and rural communities in California. Gorman and colleagues obtained different results according to the level of unit of analysis. A significant relationship

between alcohol outlets and violence did not emerge at the municipality level of New Jersey (1998) but did at the blockgroup level in Camden, New Jersey (2001).

Teh's (2008) notable study on the location restriction investigated the effect of alcohol-outlet openings and closings on crime. She used micro-level data of liquor-store addresses and crime-incident locations in Los Angeles, along with census tract-level datasets from 1990 and 2000. Examining different opening and closing times for alcohol sales, her study identified the effects of liquor availability on crime incidents that occurred around the liquor stores. According to the study, outlet openings in the low SES neighborhoods drove most increases in property and violent crimes. In contrast, outlet closings generally incurred relatively small or reverse effects on crime incidents. In addition, both low and high SES neighborhood outlets experienced crime-displacement effects after openings and closings. However, while property crimes were displaced just to the immediate vicinity of the stores, violent crimes occurred further out. The magnitude of the displacement was greater for the low SES neighborhood outlets.

Studies on the effect of a time restriction, e.g., the Sunday "Blue Law" repeals, show that the repeal is strongly associated with alcohol sales and consumption. Stehr (2007), examining alcohol sales and consumption, argues that repealing Sunday sales policies was significantly associated with increased spirits sales in the U.S. Carpenter and Eisenberg (2009) report that the repeal policies in Ontario, Canada, increased alcohol consumption on Sunday, along with inter-temporal displacement effects of alcohol consumption on Fridays and Saturdays. These results suggest that the "Blue Law" repeal is possibly associated with an increase in crime.

The empirical findings on the relationship between the "Blue Law" repeal and crime variation provides mixed evidence. Most of the "Blue Law" studies used traffic-fatalities and crashes as their outcomes. There is a sharp contrast among the New Mexico studies. One study found that repealing the "Blue Law" in New Mexico brought about an increase in Sunday fatalities (McMillan and Lapham, 2006), but another study did not find this association (Maloney and Rudbeck, 2009). With extended samples of multiple states, Stehr (2010) shows that the crime-increasing effect of the repeal was unique to New Mexico and that the other sample states that repealed their laws did not have an increase in alcohol-related fatalities. Lovenheim and Steefel (2011), using a panel data set of 15 states that repealed the "Blue Law" between 1990 and 2009, argue that the Sunday "Blue Law" had, at most, a small effect on fatal vehicle-accident rates. They note that Sunday alcohol-sales restrictions tend to have only limited benefits on public health, less than previously believed.

The "Blue Law" studies cited above dealt with only specific outcomes of traffic fatalities and crashes. With the exception of a few studies, the effects of the "Blue Law" on general crime have not yet been fully investigated. Two Swedish studies were the forerunners. Olsson and Wikstrom (1982) report that a three-month closure of the state-run liquor-store monopoly on Saturdays, which occurred in 1981, reduced public-order crimes, domestic disturbances, and assaults. Interestingly, they also reported the possibility of an inter-temporal displacement effect of the crimes from weekends to weekdays. In contrast, Norstrom and Skog (2003) reported that the gradual repeal of a

Saturday ban on the sale of alcohol in Sweden in 2000 and 2001 did not lead to an increase in assaults on Saturdays.

Heaton's study (2012) appears to be the only one that directly investigates the effects of the Sunday "Blue Law" repeal on general crime in the U.S. He employs a difference-in-difference (Sunday vs. the other days of week, and pre-repeal vs. post-repeal) and difference-in-difference-in-difference designs (including Virginia jurisdictions that were affected by the repeals in 2004 or 2008 vs. those unaffected by the repeal) with jurisdiction-level units of observation. This Virginia study found that the repeal of the Sunday off-premise liquor-package sales ban in Virginia in 2004 and 2008 led to significant increases in minor and serious alcohol-related crimes, but no increase in other crimes. It was concluded that there was no geographic or inter-temporal displacement effect.

However, it is worth noting again that the jurisdiction unit of observation may be too large to fully describe local dynamic effects on crime. Heaton (2012)'s geographical displacement conclusion tells us only that there was no evidence of crime displacement between jurisdictions; it did not necessarily address any evidence of crime movement within a jurisdiction after the repeal of Sunday "Blue Laws." Given the limited empirical evidence on the effects of time and location restrictions on the sale of alcohol on micro-level crime patterns, more empirical research is warranted.

2.2.3. Mechanisms: How Off-Premise Alcohol Sales are Connected to Crime

This paper uses the repeal of the Pennsylvania Sunday off-premise liquor sales ban as test of the effect of alcohol availability restrictions on local crime patterns. Unlike "on-premise" alcohol sales, which look to be closely associated with violent crime and public disorder due to intoxication, "off-premise" alcohol sales are often thought to be only loosely related with crime because of time gaps between off-premise alcohol purchase and its consumption. However, off-premise alcohol sales can also be connected to crime occurrence when considering the broad effect of increased alcohol availability, not just the effect of alcohol use. When the alcohol availability is increased with the opening the W&S stores on Sundays, motivated offenders and potential victims could congregate around the Sunday-open establishments and increase crime opportunities. There are also some channels in Philadelphia in which off-premise alcohol sales are directly connected to alcohol consumption, which may lead to increases in violent crime and public disorder.

The typology of "crime generator" and "crime attractor" from the crime-pattern theory are useful mechanisms for describing the "congregation effect" due to the enhanced alcohol availability (Brantingham and Brantingham, 1995). According to crime-pattern theorists, crime attractors draw motivated offenders because they provide crime opportunities such as known suitable targets. Crime generators do not draw motivated offenders per se but by increasing the use of a space may increase the number of potential offenders who encounter serendipitous opportunities (Kurland, Johnson, and Tilley, 2013). The "crime-attraction effect" captures an image of crime-distribution

concentrations that do not have a sharp change in crime volumes. On the other hand, the term "crime-generation effect" is related with a crime volume increase, not with a crime distribution change.

Sunday off-premise liquor sales are likely associated with both attraction and generation of Sunday property crime. Opening off-premise liquor stores on Sunday in commercial areas creates a large flow of population who visit the establishments and their neighborhood commercial stores. Especially when surrounding commercial stores are closed or not populated because it is Sunday, the Sunday-open off-premise liquor stores can work as a well-functioning business place where people are concentrated on Sunday (Branas et al., 2009). The large population generates serendipitous opportunities for thefts to potential offenders who just roam around the commercial areas. This crime generation effect is similarly argued as a "flow model" in Grunewald (2007).

Also, the fact that a large number of people visit the W&S stores on Sunday attracts motivated offenders to the store-located areas. If the Sunday-open W&S stores were not open on Sunday, motivated offenders would search for suitable victims in other areas with greater efforts. A congregation of suitable victims may create another congregation of motivated potential offenders. As an anecdote evidence, some real world observations in Appendix 2-B report that there may be suspicious congregations of people in areas close to the W&S stores.

In addition, there are some cases in which off-premise alcohol sales lead to alcohol consumption. One case is in the specific context of Pennsylvania. Due to the strict regulation of on-premise alcohol licenses, a number of restaurants in Pennsylvania

allow their customers to bring own alcohol beverage. When these "Bring Your Own Beverage (BYOB)" restaurants are located close to the W&S stores, people can buy wine or liquor from off-premise stores and consume them in restaurants, which results in the same alcohol use effect as that of on-premise alcohol sales.

Another case is when people buy liquor from off-premise liquor stores and consume them immediately or later around the off-premise liquor stores. Even though off-premise liquor stores are not an ideal spot for immediate alcohol consumption, people may want to consume alcohol around the stores because alcohol outlets in general may be regarded as a sign of disorganization from the social disorganization perspective (Gruenewald, 2007; Gruenewald et al., 2006). Those who consume alcohol around the stores can be persons who gather around the store-located commercial areas for small chats, standing and drinking light amounts of alcohol. Or, they can be homeless individuals who buy cheap liquor from the W&S stores with money received from begging and consume the purchased liquor on streets immediately upon purchase. Appendix 2-B describes such real world observations.

Those who consume alcohol around the stores or at the BYOB restaurants and are intoxicated may become uninhibited and want to commit public order offenses without feeling guilty. Being intoxicated, they may lose self-control and behave violently themselves, or mistakenly interpret social cues from others as violent ones and react violently. They also may be less careful when they are intoxicated than when they are sober, thus more easily falling victim to roaming criminals (Branas et al., 2009; Gruenewald, 2007; Gruenewald et al., 2006).

2.3. Study Design

2.3.1. Data Sources

Micro-level crime incident data were extracted from CompStat, the Philadelphia Police Department's crime-response management system.³² The data are limited to crime incidents occurring between January 1, 1998, and December 31, 2011 (a total of 5,113 days), within the boundary of Philadelphia County, comprising 1,086,694 incidents. Each incident observation contains XY coordinates, incident types, and exact year-day-hour times of the police service call or report. Among the collected data, 15,438 incidents (1.42%) were removed from the current data set due to a lack of XY coordinates.

Incidents are categorized into one of seven types of crimes: homicide, robbery, and aggravated assault ("violent crime"), burglary and all thefts ("property crime"), disorderly conduct and public drunkenness ("misdemeanor").³³

Data on all 94 Pennsylvania W&S stores that ever existed in Philadelphia during the 14 years from 1998 to 2011 were obtained from the PLCB through individual correspondence.³⁴ The data include information on whether and when a W&S store was allowed to open and sell liquor on Sundays, store IDs, location addresses, store openings and closings, relocations, and Sunday tax revenue information for each treatment store since 2009. In addition, tract-level median household income data were obtained from the

³² I am grateful to Dr. John M. MacDonald for providing access to the de-identified Philadelphia crime incident data, which he obtained from the Philadelphia Police Department through the University of Pennsylvania's Cartographic Modeling Lab (CML).

³³ Illegal-dumping incidents that were originally included in the data set were removed because they are highly unlikely to be correlated with alcohol use. Also, incidents that occurred between 2007 and 2009 lack the hour information.

³⁴ I am also grateful to all the correspondents in the Pennsylvania Liquor Control Board who provided the whole data related to the Philadelphia Wine & Spirit stores.

decennial 2000 census. In the current study, the census tracts of interest are limited to those in which any W&S store was ever located during the 14 years.

2.3.2. Samples and Variables

Confounding physical and Sunday-open policy changes in Philadelphia make it hard to identify directly the effect of the Sunday "Blue Law" repeal on crime in Philadelphia. Between 1998 and 2011, some of the W&S stores in Philadelphia moved locations, other stores were closed, and yet others were newly opened. In addition, the number of Sunday-open W&S stores varied during this period. Some of the W&S stores originally granted Sunday-open permission on February 9, 2003 lost that status after 2003, while others gained that status after 2005 (see Appendix 2-A for yearly variations of the numbers of Sunday-open permission stores in Philadelphia).

One possible strategy to avoid such potential confounding is to narrow the range of the applicable W&S stores to only those originally granted Sunday-open permission in 2003 and never experienced any physical change. The current study excluded any W&S store that experienced any physical change of relocations, openings, and closings during the 14-year period because the current study confines its research window to alcohol availability variations derived from the "Blue Law" repeal, not those derived from the store physical changes. In addition, the current study also excluded any W&S store that newly gained or lost the Sunday-open permission after February 9, 2003.³⁵ Thus, among

³⁵ It may be argued that the W&S stores that did not experience any store physical change but did experience the Sunday-open permission policy change after February 9, 2003, should be included. Those late Sunday-open permission stores were mainly granted the permission in 2005 and later, after the

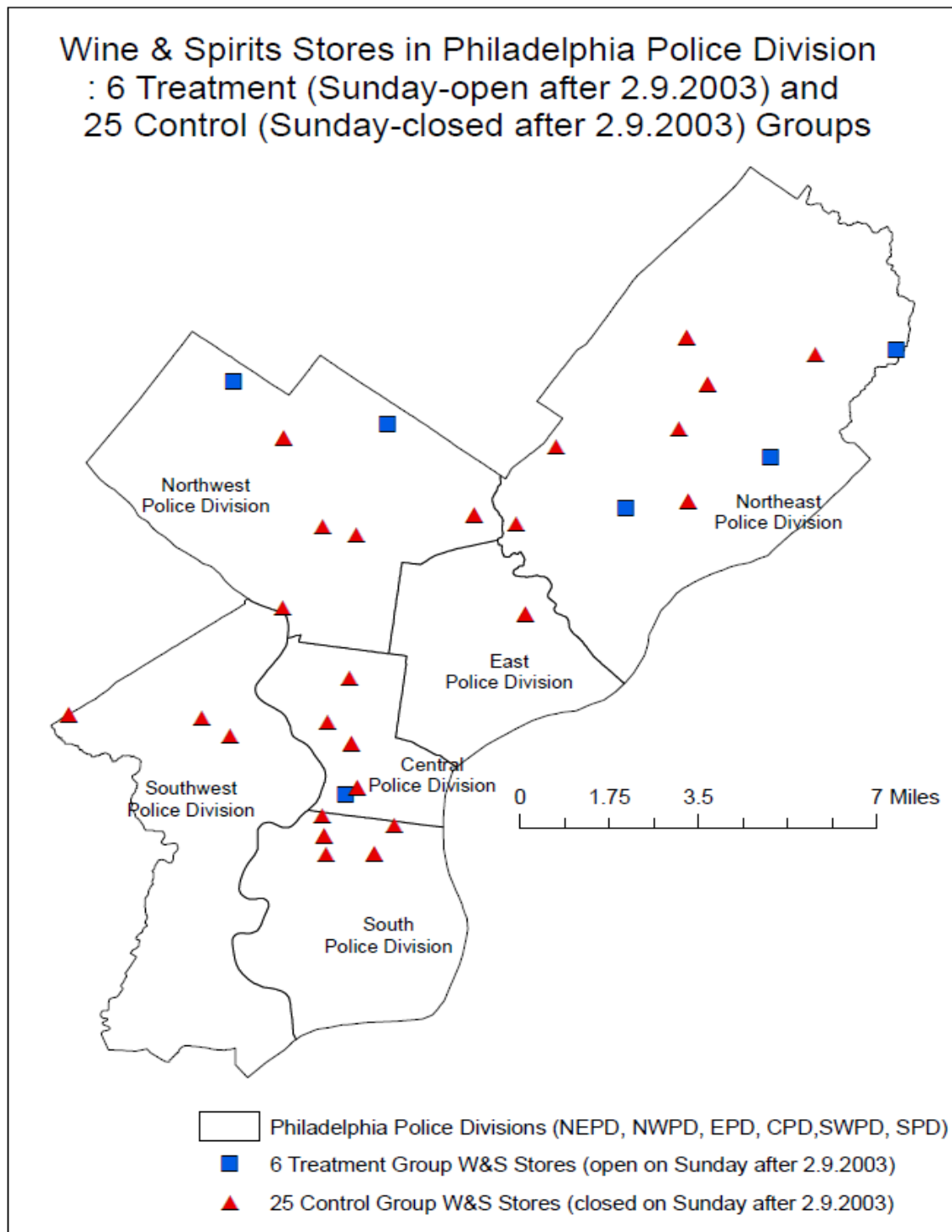
the total of 94 W&S stores that ever existed in Philadelphia for the 14 years from January 1, 1998 to December 31, 2011, in total 63 were excluded from the current sample.

Therefore, the final sample of the current study consists of the 31 Philadelphia W&S stores that were continuously open on the same location during the 14 years (hereafter the entire sample). Among the 31 W&S stores, six stores acquired permission to open on Sundays on February 9, 2003 (hereafter the treatment group). The other 25 W&S stores were never allowed to open on Sundays during the period (hereafter the control group). All the 31 stores are depicted in Figure 2.1, along with a 1/8 mile radius around each site.

In addition, the entire sample stores were split into two groups, high and low socioeconomic status (SES) neighborhoods. The neighborhood unit in the current study was defined as the census tracts from the census 2000 in which any W&S store was located. The current study used an inflation-adjusted median household income of \$50,110 in the 2011 dollars as the distinction criterion between the high and low SES, corresponding to the nationwide 2011 median household income of \$50,110 (DeNavas-Walt, Proctor, and Smith, 2013). Among the 31 entire samples stores, 12 stores (4 Sunday-open and 8 Sunday-closed stores) were located in the high SES neighborhoods, while 19 stores (2 Sunday-open and 17 Sunday-closed stores) were located in the low SES neighborhoods.

number of Sunday-open W&S stores was expanded as a result of an additional amendment to the Liquor Code in 2004. However, when those late Sunday-open stores were included in the treatment group, the pre- vs. post-repeal difference cannot be uniquely identified for the control group because of the different timings of Sunday-open permission endowments.

Figure 2-1. The Locations of the 6 Treatment and 25 Control Group W&S stores in Philadelphia



The unit of observation in the current study reflected locality and time simultaneously: a 1/8 mile radius area of each W&S store and a day. Note that a day in the current study runs from 12:00:00 AM to 11:59:59 PM. The 1/8-mile distance — equivalent to a three-minute walking distance or about 1.5 city blocks — is a common unit measure employed in urban-planning research (for example, see Department of Transportation of City of Sacramento, 2006). In addition to the basic 1/8 mile radius locality, the current study also considered a variety of expanded radii areas — within 1/4 mile, within 1/2 mile, between 1/8 and 1/4 mile, and between 1/4 and 1/2 mile radius areas from each of the 31 W&S stores. A "between X and Y mile-radius area" indicates a doughnut-shaped area that sits at the center of the X-mile radius circle.

The dependent variable is established based on the unit of observation -- the number of crime incidents occurring within a relevant radius of a W&S store on an individual day. However, because the numbers of individual crime incidents occurring within the relevant mile radii of the W&S stores are relatively low,³⁶ the incidents numbers were aggregated into total, violent, property crimes, and misdemeanor categories.

Areas located within overlapping layers of the large mile radius areas create difficulties in counting numbers of crime incidents. Fortunately, there are no overlapping areas for the 1/8 mile radius areas around the 31 W&S stores. However, some of 1/4 mile and 1/2 mile radii areas do overlap, and a crime incident can occur in the overlapping

³⁶ The only exception is the crime of theft ("All Thefts" in the current data set). As an individual crime category, theft had noticeably high numbers of incidents that occurred within the relevant mile radii of the W&S stores. The DDD estimate for this category is reported in the robustness check in Table 6.

area. When there were area overlaps, the current study assigned the crime incidents located in the overlapped areas to the closest stores.

The independent variables of main interest were a dichotomous variable indicating the double-interaction term in the DD models and the triple-difference interaction term in the DDD model. For the latter, the variable was set as one only if a day fell on or after the repeal date (February 9, 2003); the day was a Sunday; and the W&S store was allowed to open and sell liquor on Sundays. For the former term, the variable was set as one if two of the three conditions were met. Otherwise, the dichotomous variable was assigned to zero.

Geographical and temporal variables were included in the regressions to partially adjust for possible confounding effects. Including those control variables reduces standard errors of coefficients, enhancing the statistical power of detecting significant effects. Both month and year fixed effects were included to control for monthly and yearly changes that were common to all stores. Although it might be theoretically ideal to include dichotomous variables of each day as a set of fixed-effect indicators, including more than 5,000 day dichotomous variables in a regression would be undesirable in practice because of incidental parameters problems (Heaton, 2012).

Holidays were also included as dichotomous control variables in the current regressions. If people adjusted their behaviors and consumed more alcohol than usual on a holiday and/or the night before the holiday, the “holiday” effect might artificially inflate the risk of crime. To address this holiday confounding, the current study assigned the holiday variable to be one if a day fell on a holiday or its eve, as long as the eve was not

on Sunday, and to be zero otherwise. Following the definitions included under Pennsylvania Statue Title 47 (§1-102), the holidays included in the current study are New Year's Day, Dr. Martin Luther King Jr. Day, President's Day, Memorial Day, Independence Day, Labor Day, Thanksgiving Day, and Christmas Day.

Crime incidents have different distributions across the distinct areas and neighborhoods in Philadelphia (Branas et al., 2011). To control for the regional differences within Philadelphia, this study uses the six Philadelphia Police Department divisions — Northwest, Northeast, Central, South, Southwest, and East — as region controls. These 6 divisions were expected to capture different characteristics among the Philadelphia regions that are time-invariant. The police divisions were the broadest clustering unit used by the department, and each district's patrol strategy within the given division should have common traits. Thus, using a police division as a proxy for a region provides an additional advantage of controlling for differences in police-patrol activity across regions, at least to a certain degree.³⁷

2.3.3. Statistical Analysis Models

Cross-sectional time-series Poisson regression models were specified for this study. In particular, two DD models and one DDD model were constructed. Any geographical crime displacement/attraction effect were estimated based on the DDD

³⁷ Changes in police-patrol strategies at a district level typically result in changes in inputs of police-prevention activities and changes in numbers of arrests, influencing crime-incident outcomes. To address this police-patrol strategy change effect on crime outcomes, it might be ideal to use the police districts as the regional variables. However, there were frequent changes in sectioning police districts in Philadelphia during the period from 1998 and 2011. For consistency, the police division was used in the current paper instead.

model, which is of the most interest in the current paper, since the DDD model is more rigorous approach than the DD model in examining whether crime patterns were altered when stores started to sell liquor on Sundays.

A DD method was employed to overcome problems associated with a single difference model. A simple pre-post comparison model that compares crime outcomes before and after a time of Sunday-open permission is the easiest model. However, such a simple model is unlikely to address potential threats to the internal validity. For example, a major historical event coinciding with the Sunday-sales-ban repeal by chance might be a true cause of changes in outcomes. A DD model yields more reliable estimates by including an additional difference — when two comparable groups across which any confounding factor influences equally are compared, the equal effects by the confounding factor are differenced away.

The current paper suggested two sets of the DD model. The first DD model had the difference between Sundays and the other days of the week for the treatment group in addition to the pre-post comparison surrounding the repeal of the ban. In other words, the two differences were (1) whether changes in crime rates within 1/8 mile radius areas of the treatment group stores occurred on Sundays or on the other days of the week, and (2) whether the incident occurred before or after the time the treatment-group stores were allowed to sell on Sundays, February 9, 2003. Because only the treatment group stores were considered in this first DD model, the number of observations for this model was 30,678 (6 stores * 5,113 days). The effects of any confounding factor that influenced equally on Sundays and the other days of the week were expected to be differenced away.

This DD modeling is especially useful in adjusting for local confounding factors at a micro-level, such as store-specific changes.

The first DD model can be specified in a regression form. Model 1 below represents the first DD model, with the ***Sun*Post_{ijt}*** main interaction term variable, which were assigned to one if a day falls on a Sunday after the repeal, and to zero otherwise. In the model, ***i*** stands for each of the six treatment W&S stores (i=1, 2..., 6), while ***j*** stands for a day of week (j=Sun, Mon, Tue..., Sat) and ***t*** for one of 5,113 dates from 1998 to 2011 (t=1/1/1998..., 12/31/2011). ***Y_{ijt}*** indicates a number of crime incidents occurring within the relevant mile radius as a response variable. A dichotomous variable, ***Sun_j***, is assigned to one if the day of the week is Sunday, and to zero otherwise. ***Post_t*** is assigned to one if the date falls during the pre-repeal period, and to zero if it falls during the post-repeal period. ***FE_Year/Month*** variables stand for the fixed effects of years and months. The five ***Police_Division*** variables are respectively assigned to one if a W&S store is located in the Northeast, Northwest, South, Southwest, or East police divisions, with the Central police division serving as the reference. ***Holiday_t*** is assigned to one if a date falls on any one of eight holidays or their eves but does not fall on Sunday, and to zero otherwise.

$$Y_{ijt} = \alpha_0 + \alpha_1 \cdot \text{Sun} \cdot \text{Post}_{ijt} + \alpha_2 \cdot \text{Sun}_{ij} + \alpha_3 \cdot \text{Post}_{it} + \text{FE_Year} + \text{FE_Month} + \text{Police_Division1} + \dots + \text{Police_Division5} + \text{Holiday}_t + \varepsilon_{ijt} \dots \dots \dots (\text{Model 1})$$

Another DD model exploited an additional difference of Sunday crime incidents between the treatment and control groups, looking instead at the difference in crime

incidents for the treatment group only between Sundays and non-Sundays. This second model compared crime incidents occurring on Sundays only (1) between the treatment group and the control group among the entire sample, in addition to (2) the comparison between before and after the repeal of the Sunday liquor sales ban on February 9, 2003. Because this second DD model counted crime incidents occurring only on the 730 Sundays occurring between January 1, 1998, and December 31, 2011, the number of observation was 22,630 (31 stores * 730 days). Modeling with these two differences renders an advantage in that any effect of a confounding factor on crime outcome is differenced away, as long as the treatment and control groups are affected equally by that factor. This modeling is especially useful in addressing global confounding factors at a macro-level, including shifts in overall crime trends or the effect of the economic recession on crime.

The second DD model can also be specified in a regression form, as depicted in Model 2 below. Notations are in general identical to those in the first model but with a few differences. The main interaction term variable in this model, *Treat*Post_{it}*, is assigned to one if a W&S store is allowed to open on Sunday by the repeal and the date is on or after February 9, 2003, and to zero otherwise. Now *i* covers all the 31 W&S stores (*i*=1, 2..., 31), but *t* is limited to 730 Sundays occurring between January 1, 1998, and December 31, 2011. Also note that there is no *j* term in this model. *Treat* is assigned to one if a W&S store belongs to the treatment group, and to zero otherwise.

$$Y_{it} = \beta_0 + \beta_1 \cdot \textit{Treat*Post}_{it} + \beta_2 \cdot \textit{Treat}_i + \beta_3 \cdot \textit{Post}_t + \textit{FE_Year} + \textit{FE_Month} + \textit{Police_Division1} + \sim + \textit{Police_Division5} + \textit{Holiday}_t + \varepsilon_{it} \dots\dots\dots (\textit{Model 2})$$

The two DD models, however, have their own weaknesses. The strength of one can be regarded as the weakness of the other. The first model — with the Sunday vs. non-Sunday and the pre- vs. post-repeal differences — is easily influenced by macro-level changes, such as an effect of an economic boom or recession. On the other hand, the DD model — including the treatment vs. control groups and the pre- vs. post-repeal differences — cannot appropriately respond to particular local variations.

Adding another (the third) difference to a DD model can address the limitations of the DD models. The DDD model has three differences: (1) whether a crime incident occurring within the 1/8 mile radius areas of the treatment group occurred on a Sunday or on the other days of the week, (2) whether a W&S store belonged to the treatment group or the control group, and (3) whether the incident occurred before or after February 9, 2003, when the six treatment-group stores were allowed to open and sell liquor on Sundays. This DDD model is expected to yield the most reliable estimation results on the effects of the repeal on crimes occurring within the 1/8 mile radius areas of the W&S stores in Philadelphia.

The regression form of the DDD model is depicted in Model 3 below. Notations are in general identical to those in the previous DD models, except that now i stands for each of the 31 W&S stores ($i=1, 2, \dots, 31$) and t for each of 5,113 dates from 1998 to 2011 ($t=1/1/1998 \dots, 12/31/2011$). Note also that the variable of main interest in this DDD model is $Treat \cdot Sun \cdot Post_{ijt}$, that is assigned to one only if a W&S store belongs to the treatment group, the day falls on Sunday, and that a date is on or after the repeal; all other values are assigned to zero.

$$\begin{aligned}
Y_{ijt} = & \gamma_0 + \gamma_1 \cdot \textit{Treat} \cdot \textit{Sun} \cdot \textit{Post}_{ijt} + \gamma_2 \cdot \textit{Treat} \cdot \textit{Sun}_{ij} + \gamma_3 \cdot \textit{Treat} \cdot \textit{Post}_{it} + \gamma_4 \cdot \textit{Sun} \cdot \textit{Post}_{jt} + \\
& \gamma_5 \cdot \textit{Treat}_i + \gamma_6 \cdot \textit{Post}_t + \gamma_7 \cdot \textit{Sun}_j + \textit{FE_Year} + \textit{FE_Month} + \textit{Police_Division1} \\
& + \sim + \textit{Police_Division5} + \textit{Holiday}_t + \varepsilon_{ijt} \dots\dots\dots (\textit{Model 3})
\end{aligned}$$

2.4. Results

2.4.1. Descriptive Statistics

The upper half of Table 2.1 provides the overall distribution of crimes that occurred across Philadelphia for the 14 years from 1998 to 2011. Overall, more than a million incidents occurred across the city during that period. One-quarter were violent crimes, while about two-thirds were property offenses. The "All Thefts" property-crime category comprised half the total crime incidents. Incidents occurring on Sundays made up nearly one-seventh of the total.

The overall crime distribution pattern largely holds for a subset of interest — crime incidents occurring within the 1/8 mile radius areas of the entire sample (the lower half of Table 2.1). This subset covered only 3.2% of the total crime incidents. Property-crime incidents were still dominant, and the overall share of Sunday crime incidents was around 12%. The proportion of Sunday property crime incidents over the total crime incidents for these subset areas was higher than that for the entire city — 74.7%, compared to 68.5%. Instead, Sunday violent-crime incidents for the subset areas occupied a smaller share of the Sunday total, 22.1% compared to 28.8%.

Table 2-1. Types and Numbers of Crime Incidents, According to Days of the Week

	On all days of the week	On Sundays	Sunday shares of crime incidents
<i>Crime incidents occurring across the entire City of Philadelphia</i>			
Total Crime	1,071,256 [100%]	139,647 [100%]	13.0%
Violent Crime	267,299 [25.0%]	40,278 [28.8%]	15.1%
Homicide	5,644	956	16.9%
Aggravated Assault	129,612	21,051	16.2%
Robbery	132,043	18,271	13.8%
Property Crime	733,573 [68.5%]	88,927 [63.7%]	12.1%
Burglary	159,351	18,961	11.9%
All Thefts	574,222	69,966	12.2%
Misdemeanor	70,384 [6.6%]	10,442 [7.5%]	14.3%
Disorderly Conduct	65,343	9,735	14.9%
Public Drunkenness	5,041	707	14.0%
<i>Crime incidents occurring within the 1/8 mile radius areas of the 31 W&S stores</i>			
Total Crime	34,038 [100%]	4,221 [100%]	12.4%
Violent Crime	6,572 [19.3%]	932 [22.1%]	14.2%
Homicide	114	18	16.0%
Aggravated Assault	2,700	448	16.6%
Robbery	3,758	466	12.4%
Property Crime	25,429 [74.7%]	2,978 [70.6%]	11.7%
Burglary	3,753	479	12.8%
All Thefts	21,676	2,499	11.5%
Misdemeanor	2,037 [6.0%]	311 [7.4%]	15.3%
Disorderly Conduct	1,731	282	16.3%
Public Drunkenness	306	29	9.5%

Note: Only 3.2% of the total crime incidents occurring across the Philadelphia city (=33,830/1,071,256) occurred within the 1/8 mile radius areas of the 31 interested W&S stores. Percentages in brackets indicate shares of given crime categories' incident numbers over the total crime incident numbers.

Given the 158,503 observation sample size for the DDD model (=5,113 days*31 stores), the volumes of individual crime incidents occurring within the 1/8 mile radius areas of the W&S stores were too low. Except for "All Theft," they had only several thousand non-Sunday incidents and hundreds of Sunday incidents. These low volumes might lead to imprecise estimations. To overcome this low volume limitation, the current study's outcomes were aggregated based on offense type: violent, property, misdemeanor, and total crimes. Note again that, despite the aggregation, the volumes of misdemeanor (N=1,988) and violent crime incidents (N=6,548) were still low, thus potentially leading to unstable and relatively imprecise estimations.

Table 2.2 provides the comparisons of average daily numbers of crime incidents. Notice that the values in cells indicated raw average numbers of crime incidents occurring within a 1/8 mile radius of a W&S store for each day, without any control for confounding factors. Averages were calculated by dividing the summed numbers of relevant crime incidents for a given day of the week (Sunday or non-Sunday) during given periods (pre- or post-repeal) for given groups (treatment or control) by the relevant total numbers of days during given periods. The denominators were 1,865 days for the pre-repeal period (from January 1, 1998, to February 8, 2003) and 3,248 days for the post-repeal period (from February 9, 2003, to December 31, 2011).

Table 2-2. Average Daily Crime Incidents Occurring within the 1/8 mile radius areas of the 31 W&S stores

	Treatment Group: 6 W&S Stores allowed to open on Sundays since Feb. 9, 2003			Control Group: 25 W&S Stores not allowed to open on Sundays since Feb. 9, 2003		
	Pre-repeal (1/1/98~ 2/8/03)	Post-repeal (2/9/03~ 12/31/11)	<i>Pre vs. Post Difference</i>	Pre-repeal (1/1/98~ 2/8/03)	Post-repeal (2/9/03~ 12/31/11)	<i>Pre vs. Post Difference</i>
<i>Total Crimes</i>						
On Sundays	0.216	0.212	-0.004	0.208	0.164	-0.044
On non-Sundays	0.280	0.216	-0.064	0.247	0.196	-0.051
<i>Violent Crimes</i>						
On Sundays	0.037	0.031	-0.006	0.043	0.043	.
On non-Sundays	0.038	0.032	-0.006	0.045	0.042	-0.003
<i>Property Crimes</i>						
On Sundays	0.178	0.154	-0.024	0.148	0.110	-0.038
On non-Sundays	0.227	0.170	-0.057	0.188	0.142	-0.046
<i>Misdemeanors</i>						
On Sundays	0.008	0.021	+0.013	0.017	0.011	-0.006
On non-Sundays	0.015	0.014	-0.001	0.014	0.011	-0.003

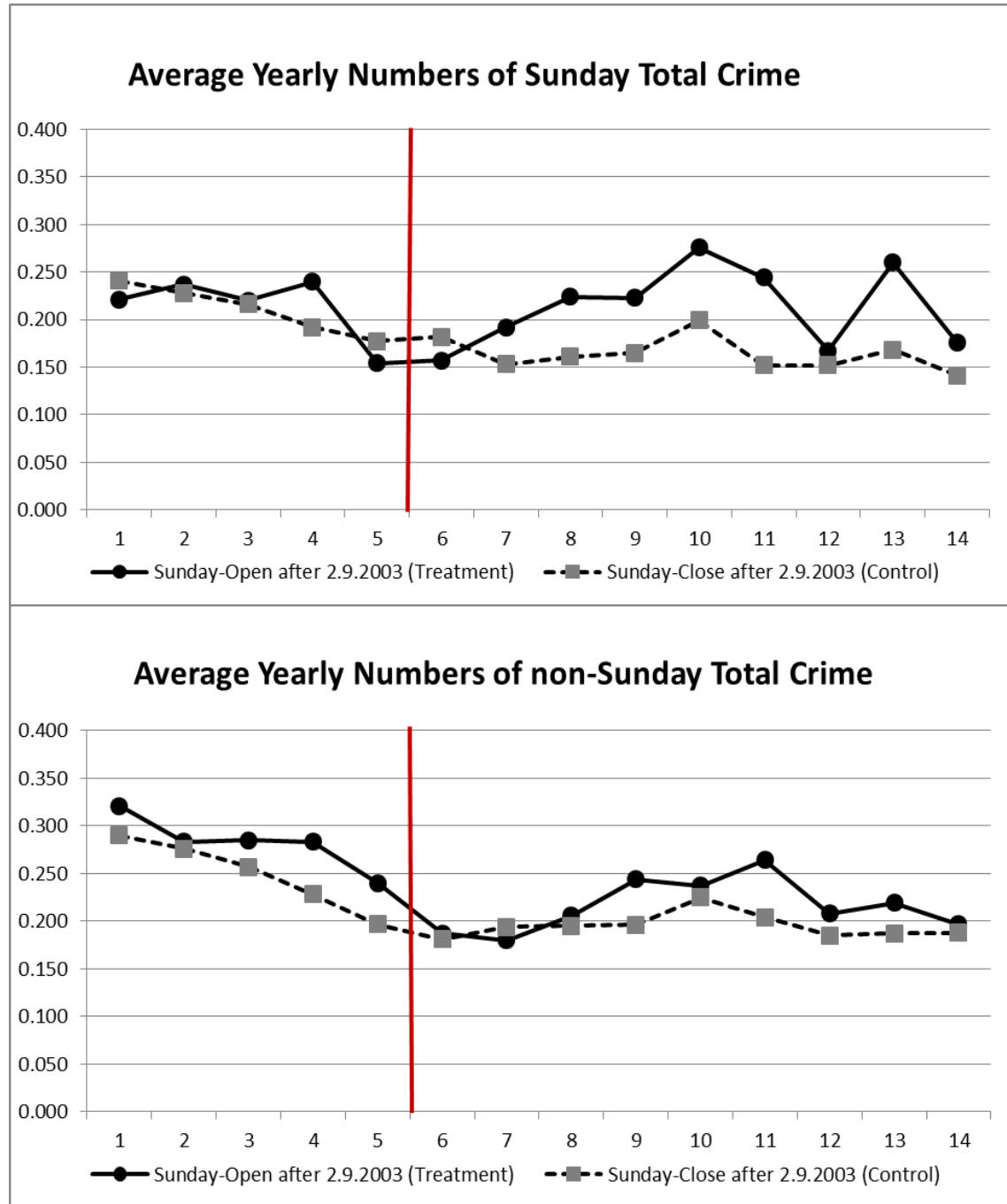
Note: The average daily numbers of crime incidents are calculated based on 5,113 days, of which the pre-repeal period was 1,865 days from 1/1/1998 to 2/8/2003 and the post-repeal period was 3,248 days from 2/9/2003 to 12/31/2011. A value in a cell represents a real average number of incidents.

Across the crime categories, differences between the pre- and post-repeal periods were noticeable. First, overall, the post-repeal period tended to have a lower volume of average daily incidents than the pre-repeal period, meaning that the general crime trend in Philadelphia was decreasing. Second, the absolute magnitudes of differences for crime incidents on non-Sundays were in general larger than those on Sundays except for misdemeanors. Third, the treatment group in general experienced large reductions in the non-Sunday crime incidents but relatively small reductions in the Sunday crime incidents after the repeal. Misdemeanor incidents even increased for the treatment group. On the other hand, the control group experienced rather similar reductions in both Sunday and

non-Sunday crime incidents. Fourth, both the raw numbers and the absolute difference magnitudes for violent crime and misdemeanor incidents seem to be too small for a statistical test. Any statistical result for violent crime and misdemeanor incidents should be cautiously interpreted because the variations might be sensitive to small changes in incident numbers.

Figure 2.2 describes the trend of average yearly numbers of total crime incidents occurring within the 1/8 mile radius areas of both the treatment and control group W&S stores. For the average yearly numbers of Sunday total crime incidents (the upper part of Figure 2.2), the differences between treatment and control groups are relatively small before the repeal, but they became large after the repeal. On the other hand, the average yearly numbers of non-Sunday total crimes (the lower part of Figure 2.2), the differences between treatment and control groups remained relatively stable even after the repeal. These two trend graphs roughly suggest the possibility that the Sunday liquor sales ban repeal in 2003 may be associated with a crime increase.

Figure 2-2. Average Yearly Total Crime Incidents Occurring within the 1/8 mile radius areas of the 31 W&S Stores



Note: The line between the years of 2002 and 2003 roughly indicates the quasi-experiment timing (repealing the Sunday liquor sales ban on February 9, 2003). The upper half depicts a trend of average yearly numbers of total crime incidents occurring on Sundays, and the lower half depicts those occurring on the other days of the week.

2.4.2. Difference-in-Difference and Difference-in-Difference-in-Difference

Results

Although Table 2.2 roughly described the crime changes before and after the repeal according to offense types and the days of the week, the raw daily average numbers in that table were unlikely to accurately describe the repeal effects because of potential confounding factors. For precise estimation, the DD and DDD models were introduced. Table 2.3 reports the results, respectively, for the two DD models and the one DDD model that were specified in the previous section, in terms of an average marginal effect. For the space limitation, the marginal effects of the other variables than the DD and DDD interactions were not reported here.³⁸

The first row denotes the estimation results of the first DD model, which had the two differences of Sunday vs. non-Sunday and pre- vs. post-repeal for the treatment group only. For this first DD model, the coefficients for total crime and misdemeanor were statistically significant at the two-tailed 1% level. The average marginal effect for total crimes was 0.057, which implies the repeal was statistically significantly associated with a 0.057 unit increase in expected Sunday total-crime incidents for a 1/8 mile radius area per Sunday, holding all other values equal. The average marginal effect for misdemeanors was 0.016, implying that the repeal was associated with a 0.016 unit increase in expected misdemeanor incidents for a 1/8 mile radius area per Sunday. Meanwhile, the average marginal effects for violent and property crimes were 0.012 and 0.027, respectively, but they were not statistically significant.

³⁸ Coefficients for other cases are available on request.

The results of the second DD model, which had the two differences of treatment vs. control and pre- vs. post-repeal for incidents occurring on Sundays only, are reported in the second row. This second DD model had similar results as those of the first DD model. For total crime, the repeal was statistically significantly associated with a 0.041 unit increase in expected total crime incidents at the two-tailed 0.1% level, holding all other values equal. Misdemeanor incidents increased after the repeal, having the average marginal effect of 0.020. Also, the average marginal effects for violent and property crime incidents were not statistically significant.

The DDD model results, which are of main interest, are reported in the third row in Table 2.3. The DDD estimates in general resemble those in the above two DD models. The average marginal effects for violent and property crimes were also statistically insignificant. For total crime and misdemeanor, the average marginal effects remained statistically significant at the two-tailed 1% level. The average marginal effect for total crime was 0.053, being almost the same to that of the first DD model. Given that a year typically has 52 weeks and Sundays, one Sunday-open W&S store's 1/8 mile radius area is expected to experience additional 2.76 total crime incidents a year due to the repeal. The absolute value of 2.76 per se is small, although the value only represents the tiny geographical area of a 1/8 mile radius. Misdemeanor incidents had the average marginal effect of 0.017, and it corresponds to additional 0.88 misdemeanor incidents a year due to the repeal.

Table 2-3. Poisson Regression Average Marginal Effects in the DD and DDD Models

	Total Crime	Violent Crime	Property Crime	Misdemeanor
(1) α_1 in DD Model 1 (Diff: Sun/Post) (N= 30,678)	0.057** (0.018)	0.012 (0.007)	0.027 (0.016)	0.016** (0.005)
(2) β_1 in DD Model 2 (Diff: Treat/Post) (N= 21,900)	0.041** (0.016)	0.008 (0.009)	0.020 (0.013)	0.020*** (0.005)
(3) γ_1 in DDD Model 3 (Diff: Sun/Treat/Post) (N= 158,503)	0.053** (0.019)	0.013 (0.009)	0.026 (0.016)	0.017** (0.005)

Note: The unit of observation is the average daily number of relevant crime incidents occurring within one W&S store's 1/8 mile radius. The difference name "Sun" stands for the difference between crime incidents on Sunday and those on non-Sundays; "Treat" stands for the difference between the treatment and control groups; and "Post" stands for the difference between the pre-repeal and post-repeal periods. Models 1 and 2 are estimated from the DD specifications, while Model 3 is estimated from the DDD specification. Coefficients in cells are average marginal effects of predicted number of relevant crime incidents, coming from unique Poisson regressions. The robust standard errors are provided in parentheses. The entire coefficients in the models are available on request. For the statistical significance, * : $p < 0.05$, ** : $p < 0.01$; *** : $p < 0.001$.

2.4.3. The Geographical Displacement/Attraction Effect³⁹

Although Table 2.3 consistently suggested statistically significant increases in total crime and misdemeanor incidents occurring within the 1/8 mile radius areas of the treatment group stores on Sunday after the repeal, it did not address a policy-relevant question of whether an increase in alcohol availability produces a net increase in crime in

³⁹ In addition to this geographical displacement, the inter-temporal displacement effect is also investigated. In addition, the triple interaction terms for Saturdays and Mondays are included in the Sunday DDD model for this investigation. In general, no inter-temporal displacement effect on Saturdays and Mondays is detected at all. The full results for the inter-temporal displacement are available upon request.

the city. If the reported crime increases reflected those that would have occurred in other places had it not been for the repeal, these increases were just displaced from the other places to the store areas, hardly influencing the overall volume of crime incidents after the repeal.

To address this displacement issue, the current paper investigated whether there was a geographical pattern change or redistribution of crime incidents, not the exact amount of crime increases. The geographical displacement hypothesis was investigated by expanding the boundary radius around the W&S stores to 1/4 and 1/2 miles. Table 2.4 reports the crime incident pattern variations for the expanded radii areas of the W&S stores. For the comparison purpose, the first row of the table repeats the DDD estimates for crime incidents occurring within the 1/8 mile radius areas of the W&S store, which were reported in the third row of Table 2.3.

The second and third rows of Table 2.4 report the DDD estimates for crime incidents occurring within broader radii of the stores. With the unit radius expanded to 1/4 and 1/2 miles, the magnitudes of the average marginal effects were changed: from 0.053 to 0.078 and to -0.022 for total crime; from 0.013 to 0.014 to -0.026 for violent crime; from 0.026 to 0.038 to -0.037 for property crime; and from 0.017 to 0.030 to 0.038 for misdemeanor. Specifically, when the unit radius was expended to 1/4 miles, the average marginal effect magnitudes for all crime categories increased. However, with the unit radius expanded to 1/2 miles, the magnitudes of the average marginal effects for total, violent, and property crime decreased, even having negative signs, with no longer holding statistical significance. In general, expanding the radius to 1/4 and 1/2 miles led

to smaller magnitudes of the average marginal effects and lower-level statistical significances of coefficients, compared to the coefficients when the radius was a 1/8 mile.

The dramatic changes in magnitudes and signs of coefficients for the 1/2 mile radius area case led to a suspicion of particular geographical patterns — that crime might be "attracted" from farther away areas to the immediate vicinity of the Sunday-open liquor stores after the repeal. Table 2.4's fourth and fifth rows report the average marginal effects for crime incidents occurring in the doughnut-shaped areas located in between 1/8 and 1/4 miles and between 1/4 and 1/2 miles radii of the W&S stores, respectively, which tests the geographical crime-attraction effect. As the fourth row describes, there was no statistically significant change in crime incidents occurring in between 1/8 and 1/4 miles radii of the treatment group stores.

The fifth row in the table, however, describes that there were statistically significant decreases in total and property crime incidents occurring in the farther away area of the stores, delineated as an area in between 1/4 and 1/2 miles radii of the treatment stores. Total crime had the average marginal effect of -0.113, which implies that the repeal was associated with a 0.113 unit decrease in expected total crime incidents occurring in the farther away area of the stores, holding all other values equal. The average marginal effect for property crime was -0.079, implying that the repeal was associated with a 0.079 unit decrease in expected property crime incidents occurring in the farther away area of the stores, all other values being equal.

On the other hand, the average marginal effects of the repeal for violent crime and misdemeanor in the farther away area were not statistically significant. Interestingly, no

evidence was found for the geographical displacement/attraction of misdemeanor incidents, with an almost zero coefficient of -0.001 and no statistical significance. It suggests the possibility that the misdemeanor increase within the 1/8 mile radius areas of the stores might be a net increase.

Table 2-4. Geographical Displacement/Attraction Effect: DDD Estimates for Crime Incidents Occurring Within Extended Radii of the W&S Stores (N=153,390)

	Total Crime	Violent Crime	Property Crime	Misdemeanor
(1) Within 1/8 mile radius	0.053** (0.019)	0.013 (0.009)	0.026 (0.016)	0.017** (0.005)
(2) Within 1/4 mile radius	0.078* (0.034)	0.014 (0.016)	0.038 (0.027)	0.030** (0.009)
(3) Within 1/2 mile radius	-0.022 (0.056)	-0.026 (0.029)	-0.037 (0.045)	0.038 (0.016)
(4) Between 1/8 and 1/4 mile radii	0.022 (0.026)	-0.003 (0.014)	0.011 (0.021)	0.011 (0.007)
(5) Between 1/4 and 1/2 mile radii	-0.113** (0.043)	-0.044 (0.024)	-0.079* (0.034)	-0.001 (0.014)

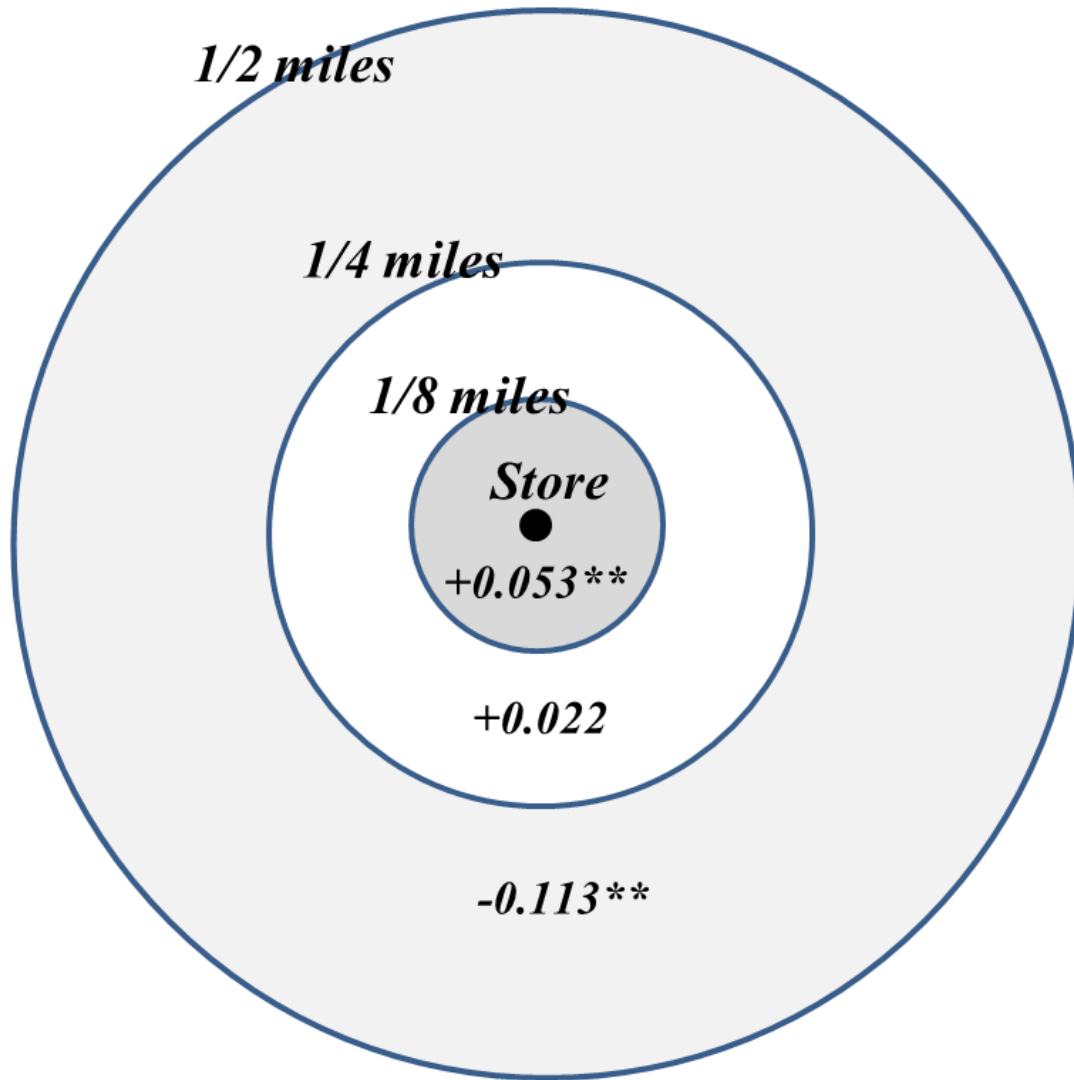
Note: The unit of observation is the average number of relevant crime incidents per one day per one W&S store's relevant mile radius. The "between 1/8 and 1/4 mile radii" refers to the strip-shaped 1/4 mile radius areas that hollows 1/8 mile radius areas within them. In the same vein, the "between 1/4 and 1/2 mile radii" refers to the strip-shaped 1/2 mile radius areas that hollows 1/4 mile radius areas within them. Coefficients in cells are average marginal effects of predicted number of events, coming from unique Poisson regressions. They are estimated from the DDD models which have the three differences between crime incidents on Sundays and those on non-Sundays, between the treatment and control groups, and between the pre-repeal and post-repeal periods. Robust standard errors are provided in parentheses. The entire Poisson regression coefficients for each model are not reported due to space limit but available on request. For the statistical significance, * : p<0.05, ** : p<0.01; *** : p<0.001.

Given all the above results, it would be plausible to say that some evidence was found for the local displacement or "attraction" effect of the repeal at least for total crime. The increase in total crime incidents occurring within the 1/8 mile radius areas of the treatment group stores after the repeal seems to be attributed, at least in part, to those that moved from farther away areas, which was in the current paper the doughnut-shaped areas located in between 1/4 and 1/2 mile radii of the treatment group stores. In other words, total crime incidents might be geographically "attracted" to the areas close to the W&S stores that were open on Sunday after the repeal. The displacement/attraction effect of total crime is depicted graphically in Figure 2.3.

2.4.4. Neighborhood Socioeconomic Status (SES) and the Displacement or Attraction Effect

The previous section demonstrated that the repeal was associated with an increase in total crime incidents occurring in the areas close to the treatment group W&S stores (delineated as those occurring within the 1/8 mile radius areas) but with a decrease in total crime incidents occurring in distant areas of the stores (delineated as those occurring in between 1/4 and 1/2 mile radii). However, the neighborhood's SES may moderate the relationship between alcohol availability and crime (Teh, 2008).

Figure 2-3. Average Marginal Effect Changes of Total Crime Incidents for Various Radii Areas



Note: The center point stands for a W&S store. The smallest circle around the center point indicates the 1/8 mile radius of the store, while the middle circle indicates the 1/4 mile radius and the largest one indicates the 1/2 mile radius. The *s next to the average marginal effects indicate the corresponding statistical significances, *: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$.

Table 2.5 reports two separate results for the geographical displacement/attraction effects in the high and low SES tract-level neighborhoods, respectively. There were noticeable differences in the crime patterns between the two groups. In general, the high SES neighborhoods (the upper half) bore no displacement/attraction pattern like that in Table 2.4. After the repeal, no statistically significant increase or decrease was expected for the all types of crime occurring within the 1/8 mile radius areas and in between 1/4 and 1/2 mile radii areas of the W&S stores. The only exception was the statistically significant increase in property crime occurring within the 1/8 mile radius areas of the stores. It is notable that the repeal was rather significantly associated with decreases in violent crime incidents occurring in distant areas of the stores, implying that the high SES neighborhoods enjoyed advantages of violent crime reduction after the repeal.

On the other hand, the low SES neighborhoods held the crime displacement or attraction patterns (the lower half of Table 2.5). Total crime incidents were expected to statistically significantly increase within the 1/8 mile radius areas of the W&S stores after the repeal, but to decrease in between the 1/4 and 1/2 mile radii areas of the stores. Misdemeanor incidents were also expected to statistically significantly increase, without any evidence of a displacement/attraction effect. It is worth noting that violent crime incidents were expected to increase within the 1/8 and 1/4 mile radius areas of the stores in the low SES neighborhoods.

Table 2-5. Geographical Displacement/Attraction Patterns for Low vs. High SES Neighborhoods

<i>High SES Neighborhoods (N=61,356) (4 Treatment and 8 Control Groups)</i>				
	Total Crime	Violent Crime	Property Crime	Mis-demeanor
(1) Within 1/8 mile radius	0.035 (0.025)	-0.009 (0.009)	0.052* (0.021)	-0.016 (0.012)
(2) Within 1/4 mile radius	-0.010 (0.037)	-0.036* (0.014)	0.043 (0.032)	-0.030* (0.012)
(3) Within 1/2 mile radius	-0.048 (0.056)	-0.046* (0.022)	0.013 (0.047)	-0.029 (0.016)
(4) Between 1/8 and 1/4 mile radii	-0.034 (0.027)	-0.026* (0.010)	-0.002 (0.023)	-0.013 (0.008)
(5) Between 1/4 and 1/2 mile radii	-0.039 (0.040)	-0.009 (0.017)	-0.028 (0.034)	-0.005 (0.010)
<i>Low SES Neighborhoods (N=97,147) (2 Treatment and 17 Control Groups)</i>				
	Total Crime	Violent Crime	Property Crime	Mis-demeanor
(1) Within 1/8 mile radius	0.062* (0.027)	0.030* (0.014)	0.005 (0.023)	0.030*** (0.007)
(2) Within 1/4 mile radius	0.152** (0.051)	0.071** (0.027)	0.034 (0.041)	0.060*** (0.013)
(3) Within 1/2 mile radius	0.035 (0.088)	0.011 (0.048)	-0.061 (0.070)	0.091*** (0.025)
(4) Between 1/8 and 1/4 mile radii	0.089* (0.041)	0.037 (0.024)	0.031 (0.033)	0.027* (0.011)
(5) Between 1/4 and 1/2 mile radii	-0.161* (0.070)	-0.072 (0.041)	-0.109* (0.055)	0.005 (0.026)

Note: Coefficients in cells are average marginal effects of predicted number of events, coming from unique Poisson regressions. They are estimated from the DDD models which have the three differences between crime incidents on Sundays and those on non-Sundays, between the treatment and control groups, and between the pre-repeal and post-repeal periods. Robust standard errors are provided in parentheses. Percentage increases are also provided in brackets. The entire coefficients sets for each model are not reported due to space limit but available on request. For the statistical significance, *: p<0.05, **: p<0.01; ***: p<0.001.

2.5. Robustness Check

Robustness checks were conducted, and the results are reported in Tables 2.6 and 2.7. Overall, the effects of repealing the Sunday liquor off-premise sales ban on crime were consistently supported. In the robustness tests, the total crime incidents occurring within the 1/8 mile radius areas of the treatment group W&S stores were consistently expected to increase with statistical significance, while those occurring between 1/4 and 1/2 mile radii of the stores were expected to decrease. For the convenience of comparison, the DDD estimates for the original model in Table 2.4's first row were repeated in the first row of Table 2.6.

The second row presents a falsification test. The average marginal effects were reported for the model that moved the focal day from Sunday to Tuesday. Due to the Sunday-specific nature of the current quasi-experiment, crime incidents occurring on Tuesdays were not expected to be significantly affected by the repeal. As expected, this falsification model yielded non-significant results, implying that the original model's results were not just products of chance or of a coincident trend that occurred on any day of the week.

The third row provides results when a measurement of a day was adjusted. Alcohol-related crime incidents occurring between midnight and 6:00 AM would be committed under intoxication from alcohol consumed the day before the day of interest. Thus, it may be ideal to measure a day by 24 hours from 6:00 AM to 5:59 AM the next day for the alcohol-crime relationship study (Heaton, 2012). However, the current data

set did not have the hour information for incidents occurring between 2007 and 2009.⁴⁰ A partial change of day measures for years except for 2007 to 2009 did not completely refute the repeal's effects, despite smaller magnitudes and less statistical significance for the average marginal effects. Overall, the directions of the repeal's effect still held.

Given that Center City (downtown of Philadelphia) is filled with a large and dense population and a number of possible crime-generating events, the patterns of crime incidents in this area might differ from those of the other areas. Whether the original DDD estimates were driven by the particularity of this sector was tested by recalculating the DDD estimations without the four W&S stores located in Center City (one from the treatment group and three from the control group).⁴¹ The fourth row in Table 2.6 reports the results, which in general were similar to the original estimates. Interestingly, however, there was no longer statistically significant association between the repeal and misdemeanor crime incident change, implying that the repeal effect on misdemeanor might be driven by other unaddressed factors than the repeal.

Estimates of individual crime changes may better describe the dynamics of the repeal effects than estimates of aggregated crime changes. The fifth row the average marginal effects estimated only for the "All Thefts" property crime incidents. Now the "All Thefts" property crime incidents were shown to statistically significantly increase within the immediate vicinity of the treatment group stores.

⁴⁰ In addition to this downside, it is worth noting that the hour information of crime incidents may not reflect exact crime incident times because it is marked only based on service call times to the police. Service calls can be placed prior to or after actual crime occurrences.

⁴¹ The store IDs are 9113 (the treatment group), 5119, 5122, and 5143 (the control group).

The choice of regression methods did not change the results. The sixth and seventh rows in Table 2.6 provide results obtained from the negative binomial and OLS regression modeling, respectively. The repeal was consistently statistically significantly associated with increases in expected total crime misdemeanor incidents in the immediate vicinity and with a decrease in total crime incidents in the farther away areas of the stores, holding all other values equal.

In addition, p-values of permutation tests, in which the originally observed data were randomly shuffled 1,000 times, are reported in Table 2.7. They pertained to proportions of the shuffled data's DDD regression coefficients whose absolute values were greater than the absolute values of the original data's DDD coefficient. They indicated how unusual the observed DDD coefficients were if the null hypothesis was true that there was no association between crime-incident variations and the repeal.

The results were supportive of the previous robustness test results. Evidence for the repeal effect on crime incidents occurring within the 1/8 mile radius areas of the stores was strongly supported. In the 1,000 shuffles, less than 1% of them produced unusual coefficients that were larger than the original DDD estimates for total crime (0.5%) and misdemeanor (0.3%). However, the repeal effect on the total crime incidents occurring between 1/4 and 1/2 mile radii of the stores was not statistically significant at the two-tailed 5% level.

Table 2-6. Robustness Check: Alternative DDD Estimates for Crime Incidents Occurring Within 1/8 Mile and Between 1/4 and 1/2 Mile Radii of the W&S Stores

	Within 1/8 mile Radius areas				Between 1/4 and 1/2 Mile Radii			
	Total Crime	Violent Crime	Property Crime	Mis-demeanor	Total Crime	Violent Crime	Property Crime	Mis-demeanor
(1) Original model (N=158,503)	0.053** (0.019)	0.013 (0.009)	0.026 (0.016)	0.017** (0.005)	-0.113** (0.043)	-0.044 (0.024)	-0.079* (0.034)	-0.001 (0.014)
(2) "Tuesday" as the false affected day (falsification test) (N=158,503)	-0.019 (0.012)	-0.004 (0.005)	-0.018 (0.010)	0.001 (0.003)	-0.036 (0.026)	0.015 (0.014)	-0.043* (0.020)	-0.014 (0.010)
(3) Measure a day from 6 AM to next day 6 AM <i>Including 2007-2009</i> (N=158,503) <i>Excluding 2007-2009</i> (N=124,527)	0.036 (0.020) 0.037 (0.022)	0.008 (0.009) 0.009 (0.010)	0.023 (0.017) 0.024 (0.019)	0.009 (0.005) 0.004 (0.006)	-0.121** (0.045) -0.130** (0.050)	-0.045 (0.026) -0.049 (0.029)	-0.066 (0.034) -0.081* (0.038)	-0.016 (0.015) -0.013 (0.018)
(4) Remove four W&S stores located in Center City (N=138,051)	0.061** (0.021)	0.009 (0.011)	0.045* (0.017)	0.007 (0.006)	-0.113* (0.052)	-0.059* (0.028)	-0.062 (0.040)	-0.009 (0.016)
(5) Estimate only for "All Thefts" (N=158,503)	-	-	0.030* (0.015)	-	-	-	-0.065* (0.028)	-
(6) Estimate by Negative Binomial Regressions (N=158,503)	0.054** (0.019)	0.012 (0.009)	0.028 (0.016)	0.015** (0.005)	-0.111* (0.043)	-0.044 (0.024)	-0.078* (0.034)	-0.004 (0.014)
(7) Estimate by OLS regressions (N=158,503)	0.053** (0.020)	0.010 (0.008)	0.026 (0.017)	0.018*** (0.005)	-0.102** (0.037)	-0.026 (0.016)	-0.076* (0.031)	-0.00004 (0.008)

Note: See the note in Table 2.4. The coefficients from the OLS regressions indicate the variation of raw numbers of crime incidents. The entire coefficients sets for each model are not reported due to space limit. For the statistical significance, * : p<0.05, ** : p<0.01; *** : p<0.001.

Table 2-7. Permutation Tests P-values for the Geographical Displacement Effect: DDD Estimates for Crime Incidents Occurring within Extended radii of the W&S Stores

	Total Crime	Violent Crime	Property Crime	Misdemeanor
(1) Within 1/8 mile radius	$p = 0.005^{**}$	$p = 0.068$	$p = 0.074$	$p = 0.003^{**}$
(2) Within 1/4 mile radius	$p = 0.016^{*}$	$p = 0.305$	$p = 0.110$	$p = 0.005^{**}$
(3) Within 1/2 mile radius	$p = 0.876$	$p = 0.135$	$p = 0.841$	$p = 0.111$
(4) Between 1/8 and 1/4 mile radii	$p = 0.342$	$p = 0.726$	$p = 0.406$	$p = 0.204$
(5) Between 1/4 and 1/2 mile radii	$p = 0.063$	$p = 0.015^{*}$	$p = 0.118$	$p = 0.784$

Note: Values in cells indicate proportions of the shuffled data's DDD regression coefficients whose absolute values are greater than the absolute values of the original data's DDD coefficient, which means the two-tailed p-values of the permutation tests. For the statistical significance, *: $p < 0.05$, **: $p < 0.01$; ***: $p < 0.001$.

2.6. Discussion and Conclusion

The current study consistently show that repealing the Pennsylvania Sunday off-premise liquor sales ban in 2003 was associated with a statistically significant increase in Sunday total crime incidents occurring within the 1/8 mile radius areas of the opened W&S stores. The increase appears to be driven by increases in property crime and misdemeanor incidents. However, the repeal did not statistically significantly change a violent crime pattern. Also, this paper provides some evidence of local displacement or "attraction" effect after the repeal. In particular, there was an increase in total crime incidents occurring within 1/8 mile radius of the stores after the repeal, while there was a

slight decrease in total and property crime incidents occurring between 1/4 and 1/2 mile radii of the stores after the repeal. Given this simultaneity, it can be inferred that at least some offenders were "attracted" to the immediate vicinity of the treatment group (delineated as within 1/8 mile radius), searching for an increased pool of potential suitable victims from the farther away areas of the stores (delineated as between 1/4 and 1/2 mile radius), although the immediate vicinity of the treatment group stores also had some crime increases "generated" by the repeal.⁴²

If the crime-attraction effects by the repeal truly exist, our interests in "Blue Law" repeal should not be confined only to how much crime increased after the repeal. Rather, more attention needs to be paid to the crime distribution patterns. As Barr and Pease (1990) appropriately pointed out, interventions redistribute costs of crimes and thus impact individuals' quality of life. For example, even when an intervention reduces crime incidents overall, the intervention cannot be desirable to those who are newly victimized by the intervention.

In the current case, it is worth tracing who have more burdens related with crime and who have less after the repeal. Like other ecological studies, the current study investigated the crime-redistribution effect at an aggregated-level of neighborhoods, not at the individual level, by using the distinction between high and low socioeconomic status (SES) neighborhoods (Gruenewald et al., 2006; Teh, 2008). The results from this study show that the repeal-effect patterns held only for the low SES tract-level

⁴² However, the current paper does not argue that all total crime incident increases near the stores after the repeal were just replaced by incidents that would have occurred away from the stores had there not been the repeal. In fact, the current paper does not, and cannot, provide the exact extent of the crime-attraction effect of the repeal because only limited surrounding areas of store were covered in this study.

neighborhoods, implying that the repeal effects on crime may be largely confined to poor neighborhoods, with greater burdens on those who live in the poor areas. The unintentional effect on crime redistribution of the repeal is in line with what Teh (2008)'s study found by investigating alcohol store openings and closings. To address this unintentional consequence on the crime burden redistribution, some policy responses targeting the poor neighborhood residents may be required, such as implementing an increased police-patrol strategy around the Sunday-open W&S stores in the low SES neighborhoods.

Related to the crime-attraction and crime-generation effects of the repeal, it may be worth considering whether the repeal policy can be justified from the cost-benefit analysis. Due to data limitations, however, a rigorous cost-benefit analysis for the current setting is not available. The current paper attempts to just roughly compare increased costs due to additional crime incidents occurring within the 1/8 mile radius area of the W&S stores after the repeal and increased net tax revenues from the Sunday-open W&S stores' Sunday sales. According to the PLCB's electronic sales records, the 6 Sunday-open W&S stores in the current paper's treatment group collected from Sunday liquor sales on average \$117,695 (in the 2014 dollars) of net tax revenues per store per year between 2009 and 2011. The taxes used in the calculation were an 18% state liquor tax and 6% state and local sales taxes.

Meanwhile, given 52 Sundays a year, the DDD estimates for total crime incidents occurring within the 1/8 mile radius areas of the Sunday-open W&S stores (reported in Table 2-4) suggests that total crime incidents increased by around 2.76 incidents

($=0.053*52$) per year. Based on the ratios of the average marginal effect values between misdemeanor (around 1/3) and felony violent and property crime (around 2/3) reported in Table 2-4, assume that 2 incidents were felony and 0.76 incidents were misdemeanor. Following Cohen and Piquero (2009)'s crime cost estimates, assume that a felony incident other than homicide typically costs \$18,522 in the 2014 dollar and that a misdemeanor incident costs \$571 in the 2014 dollar. If there had been no crime attraction effect and the repeal newly generated all the additional 2.76 total crime incidents, the estimated crime cost for the additional incidents occurring within the 1/8 mile radius areas of the W&S stores would be \$37,478 per store per year ($=2*\$18,522 + 0.76*\571).

Comparing the net tax revenue of \$117,695 and the estimated total crime cost of \$37,478, the cost-benefit analysis may lead to a conclusion that the repeal of the off-premise liquor sales ban provides a net financial gain to the state government. However, this is a rough comparison. In a micro-level analysis like the current study, an issue arises about an appropriate size of a unit of analysis. In the current setting, when the total crime incidents occurring within the 1/4 mile radius areas of the W&S stores is reflected, instead of those within the 1/8 mile radius areas of the stores, the estimated crime cost per store per year would increase because there would be more crime incidents counted. Also, when the crime-attraction effect is considered, the complexity increases more. Because the crime attraction changes locations but not the volume of crime, incidents that were just displaced or attracted from other places should not be counted as an additional cost. However, unfortunately, it is practically almost impossible to distinguish whether a crime incident is newly generated or just displaced.

The current study has limitations. Most notably, the results were based only on limited numbers of off-premise W&S stores in Philadelphia that met a strict set of criteria. There were other liquor-related distribution facilities in Philadelphia, including the W&S stores that were excluded from the current study and on-premise liquor-selling sites such as bars, restaurants, and pubs. These facilities might influence the current results through some unaddressed paths. However, there are some reasons to believe that the influences of these facilities might not be severe in the current setting. Overall, the estimates for total crime incident variations within the 1/8 mile radius areas of the W&S stores were unlikely to be seriously distorted, because virtually all W&S stores including the excluded W&S stores were more than 1/8 mile apart. Also, it is unlikely that the numbers and locations of on-premise bars, restaurants, and pubs in Philadelphia would have varied sufficiently in a way to strongly influence the results during the study period. Pennsylvania strictly regulates its total on-premise alcohol license numbers to a low level, and thus variations of on-premise licensees are expected to be lower than those of other states. Assuming that the enforcement level of the license regulation was sufficiently high in Philadelphia, the effects of on-premise liquor-selling facilities might be roughly differenced away in the DD and DDD designs. Nevertheless, future research that addresses all the possible variations of the alcohol availability, including openings and closings of the W&S stores, is warranted for more precise crime-effect estimation.

There are also some limitations related to the data availability and quality. The unavailability of alcohol-related and non-alcohol-related incident data prevented the study from further investigating patterns of crime-incident variations after the repeal.

Imperfect hour information for incidents occurring during 2007 to 2009 also diluted the robustness of the models. In addition, the low statistical power due to the low volumes of offenses could be problematic. For misdemeanor incidents, in particular, there were too many zero incident observations, which made the results for the misdemeanor incidents less precise.

Despite these limitations, the current study has its own strengths. In terms of methodological design, it adopts not only the difference-in-difference (DD) but also the difference-in-difference-in-difference (DDD) methodology, thus yielding more reliable estimates for the effects of the "Blue Law" repeal on crime than those of simple fixed-effect regression models. Compared to the results obtained from relatively simple regression designs, the current study's DDD estimates show a higher level of evidence. In addition, by looking at the micro-level criminal behavioral changes with spatially well-defined crime venues, this study shows the unexpected existence of a crime-attraction effect toward the stores, which also leads to the issue of a disproportionate redistribution of crime burdens between high and low SES tract-level neighborhoods. These results on the micro-level crime pattern changes are compensatory to the prior literature's results suggesting no displacement effect at a macro-level.

APPENDIX 2-A. Variations in Numbers (#) of the W&S Stores in Philadelphia

Table 2-8. Variations in Numbers of the W&S Stores in Philadelphia

	Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Change in Physical Operation	# of total stores as of Jan. 1.	77	73	71	68	65	65	65	63	62	60	59	58	57	54
	# of openings that year	0	3	0	2	2	4	1	2	0	0	1	0	2	0
	# of closings that year	4	5	3	5	2	4	3	3	2	1	2	1	5	0
Change in Sunday-open Permission	# of Sunday-open stores as of Jan. 1.	0	0	0	0	0	0	8	8	15	16	15	15	18	19
	# of new permissions that year	0	0	0	0	0	8	0	8	2	0	0	3	5	1
	# of permission repeals that year	0	0	0	0	0	0	0	1	1	1	0	0	4	0

APPENDIX 2-B. Real World Observations about W&S Stores in Philadelphia, PA

*** At a W&S store on the 326 S 5th St, Philadelphia (Center City) on May 4th, 2014**

- At around 2:05pm: An old and poor-looking person, who looked already lightly-intoxicated with something while staggering, visited the W&S store. He came out with the liquor store plastic bags and with some other black plastic bags. When he saw me standing by the store, he yelled repeatedly, "Are you waiting for a woman?" He still staggered with the bags. I didn't answer them, and he passed by me, repeatedly yelling at me, and went away very slowly into the residence area near the store. 0.2 mile away from the store, there were many restaurants, including those allowing their customers to bring own alcohol beverage (BYOB).

*** At a W&S store on the 1218 Chestnut St, Philadelphia (Center City) on May 4th, 2014**

- Between 3:05pm and 4:20pm: When I reached the store, a few people gathered one block away from the store, standing or sitting in front of an old bank building. Their ages were various, and most of them wore worn-out clothes. Some of them carried their bags and other stuffs, as if they were homeless people. One of them was wearing relatively neat old military uniforms with a black backpack. They did nothing suspicious but just were chatting with each other. From time to time, one or two people roamed away from the gathering and came back. Even new faces joined the group. Sometimes one or two

passed by the W&S store gate area, but it seemed that they did not pay much attention to the store. At around 4:00pm, they were disbanded. One of them went in the store and came out with a W&S store plastic bag. All but one disappeared then. The remaining one, looking like a homeless person, moved across the street, and sit on the street with his bag and other stuffs. He stayed there for around 15 minutes and went away.

- At around 3:20pm: An old person came to the store, drawing a cart filled with flower bunches. He looked like a person who was selling these bunches on streets. However, he looked very poor with worn-out clothes. If he had not had the flower cart, I would have thought he was a homeless person. When he came out of the store, he picked out a water bottle from his cart and poured out water from the bottle, and seemed to refill the bottle with liquor he bought from the store, as if he wanted the bottled to look like just water-filled. He did not drink it at the spot. After the work, he started to draw the cart to the City Hall direction and disappeared.

- Between 4:00pm and 5:00pm: The W&S store was to be closed at 5:00pm on Sunday. After 4:00pm, there were more visitors than before, and around 4:40pm was the peak time. Some people visited in a group, while others came alone. Some came to the store with bikes and backpacks, while some just came to the store with nothing with them as if they lived around the store or just came out of their working places near the store. Walking into the store at 4:45pm, I found that many of them (especially those young) were buying discounted liquor bottles. A few roamed around the wine corners. While some bought several bottles, a majority of them bought just one bottle. I thought that they

might consume the alcohol they bought at BYOB restaurants that day evening because otherwise, they would buy more than one bottle for the storing purpose.

- After 5:00pm: The store was closed. However, there were people who came to the store late, finding the store closed. One of them looked like a homeless person. Although he had a cart-linked bike, the cart was filled with blankets, cola bottles, and some clothes. He knocked the store door but the door guard would not open the door. The homeless-looking person waved money notes in front of the door, asking the guard to let him in. However, the guard did not open the door. The person was standing in front of the door for around 5 minutes, finding the door kept closed. Then, he organized his cart and went away.

*** At a W&S store on the 4906-09 Baltimore Ave, Philadelphia on May 11th, 2014**

- Between 4:00pm and 5:00pm: Arriving at the store-located commercial area, I found two groups gathering around the store. Just next to the W&S store gate, two persons were standing and chatting while drinking canned beverage. (I could not identify whether it was alcohol or not.) They kept standing there until I left the spot after 5:00pm. Another group, consisting of 5 persons, was standing one block away from the W&S store, in front of a closed laundry shop. They did not drink any at all. They just kept chatting loud. A new face joined the group and two of them left the group. However, when a police patrol car drove into the parking lot, the remaining four of the group slowly moved out of the commercial area and went away. It seemed that the policepersons received calls from the commercial stores. The policepersons visited the W&S store and an ethnic restaurant

next to it, and came back to the patrol car and stayed there for a while, as if they made reports on what happened there. When I approached the W&S store gate from the parking lot, I found some black plastic bags thrown away at the corner of the W&S store. In the bags, there were a few empty cans of beer and cocktail alcohol, although beer was not sold at the W&S store but at a store located in few blocks away from the W&S store.

- At around 7:00pm: I left the spot after the W&S store was closed at 5:00pm, and I returned to the place again at around 7:00pm on the same day by driving. In front of the closed laundry shop, a homeless-looking person was sitting while drinking canned beverage with a black plastic bag at his sides. (I could not identify whether it was alcohol or not.)

APPENDIX 2-C. DDD Results for the Full Set of the 94 W&S Stores

This Appendix 2-C section reports the DDD analysis results when all the 94 W&S stores that were ever established in Philadelphia between January 1st, 1998 and December 31st, 2011 were considered. The 94 W&S stores were identified depending on whether the store addresses were changed; each of the 94 stores had a different address from the others. 25 stores of them ever had the Sunday-open permission during the period, being a treatment group, while the rest 69 did not at all, being a control group. Among the 94 W&S stores, 56 stores experienced any physical change such as a store opening or closing during the 13-year period. Of the remaining 38 stores that never experienced any physical change, 7 stores had any Sunday-open permission change since the initial permission date of February 9, 2003, including the repeal of the permission and the endowment of the permission after the initial date in 2003.⁴³

The current full set DDD analysis has additional limitations, compared to the reduced set analysis in the main body. One of them is that it could not identify a pre- vs. post-period difference for the control group because of the different timings of permission endowments for the treatment that varied from February 9, 2003 to December 7th, 2011. The pre- vs. post-period difference was identified only for the treatment group according to their own permission dates. In that sense, the current full set analysis is not a complete DDD analysis; while both the pre- vs. post-period and the Sunday vs. non-Sunday differences were calculated for the treatment group, only the Sunday vs. non-

⁴³ The reduced set analysis was conducted in the main body for the 31 stores that never experienced any physical or permission change,.

Sunday difference was calculated for the control group. The current analysis seems to exist between DD and DDD analyses. Also, because the pre- vs. post difference was identified only for the treatment group, a collinearity problem arose when regressions were run between a variable indicating treatment vs. control groups and a variable indicating pre- vs. post-repeal periods. In the regressions for the following table results, two variables were omitted due to the collinearity; a dichotomous variable for the pre- vs. post-repeal period difference, and a dichotomous one for the double interaction between the treatment/control and pre/post differences.

Table 2-9 reports the DDD estimate results for crime incidents occurring within 1/8 mile and extended radii of the full 94 W&S stores. Compared to the reduced set analysis results of the 31 W&S stores (reported in Table 2-4 in the main body), the full set analysis results showed similar but less evident patterns. Although an increase in total crime incidents occurring within a 1/4 mile radius of the W&S stores was still statistically significant at the two-tailed 5% level, an increase in total crime incidents occurring within a 1/8 mile radius of the W&S stores was no longer statistically significant. There was still a decrease in total crime incidents occurring between 1/4 and 1/2 mile radii of the W&S stores, which was, however, no longer statistically significant. The magnitudes of the average marginal effects for the full 94 W&S stores were way smaller than those for the reduced 31 W&S stores. In general, the DDD effects for the full set of the 94 W&S stores were not as evident as those for the reduced set of the 31 stores.

However, when the full set of the 94 W&S stores were categorized into high and low SES neighborhood groups according to the median household income, there was a

sharp contrast in crime patterns after the repeal between the two neighborhood groups. Table 2-10 reports the DDD estimates for the two neighborhood groups. For the high SES neighborhood group, which consists of 10 ever-Sunday-open and 19 Sunday-closed W&S stores, there were consistent and statistically significant decreases in all kinds of crime incidents including violent and property ones (the upper part of the table). On the other hand, for the low SES neighborhood group with 15 ever-Sunday-open and 50 Sunday-closed W&S stores, there were consistent increases in total and property crime incidents as well as misdemeanors (the lower part of the table). The opposite-directions of the high and low SES neighborhood DDD effects cancelled out each other when the full 94 W&S stores were considered together, making the total DDD effects in Table 2-9 less evident.

In general, the current full set analysis results seem to be in line with those of the reduced set analysis in the main body. Although there were somewhat different patterns between the two analyses, the different patterns may be attributed to the incomplete DDD design of the full set analysis. Note again that the full set analysis of the 94 W&S stores could not identify the pre- vs. post-repeal difference for the control group. In addition, the physical changes of the 56 W&S stores may influence the DDD estimates of the full set analysis. For example, when a W&S store was originally located in an area with a high volume of crime in the pre-repeal period but later relocated in an area with a relatively low volume of crime in the post-repeal period, the DDD estimates would become smaller than those when there would have been no relocation of the W&S store. Therefore, the current full set analysis results needs to be interpreted with cautions. The reduced set

analysis results for the 31 W&S stores in the main body, which are free from these pitfalls, seems to be more reliable than the current full set analysis results.

Table 2-9. Geographical Displacement/Attraction Effect: DDD Estimates for Crime Incidents Occurring within Extended Radii of the Full 94 W&S Stores (N=322,499) (25 Treatment and 69 Control Groups)

	Total Crime	Violent Crime	Property Crime	Misdemeanor
(1) Within 1/8 mile radius	0.012 (0.010)	0.003 (0.004)	0.012 (0.008)	-0.002 (0.003)
(2) Within 1/4 mile radius	0.035* (0.017)	-0.001 (0.008)	0.034* (0.014)	0.00005 (0.005)
(3) Within 1/2 mile radius	0.026 (0.029)	-0.005 (0.014)	0.028 (0.023)	0.010 (0.008)
(4) Between 1/8 and 1/4 mile radii	0.022 (0.014)	-0.005 (0.006)	0.022* (0.011)	0.001 (0.004)
(5) Between 1/4 and 1/2 mile radii	-0.010 (0.022)	-0.003 (0.011)	-0.008 (0.017)	0.008 (0.006)

Note: Coefficients in cells are average marginal effects of predicted number of events, coming from unique Poisson regressions. They are estimated from the same DDD model as that for the reduced set analysis in the main body. Robust standard errors are provided in parentheses. The entire coefficients sets for each model are not reported due to space limit but available on request. For the statistical significance, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

Table 2-10. Geographical Displacement/Attraction Patterns of the Full 94 W&S Stores for Low vs. High SES Neighborhoods

<i>High SES Neighborhoods for 29 W&S Stores (N=109,136) (10 Treatment and 19 Control Groups)</i>				
	Total Crime	Violent Crime	Property Crime	Misdemeanor
(1) Within 1/8 mile radius	-0.066*** (0.013)	-0.011* (0.005)	-0.034** (0.011)	-0.016*** (0.004)
(2) Within 1/4 mile radius	-0.121*** (0.020)	-0.024** (0.008)	-0.078*** (0.017)	-0.018** (0.006)
(3) Within 1/2 mile radius	-0.192*** (0.031)	-0.046*** (0.013)	-0.114*** (0.027)	-0.029** (0.008)
(4) Between 1/8 and 1/4 mile radii	-0.058*** (0.016)	-0.011 (0.006)	-0.043** (0.013)	-0.009 (0.005)
(5) Between 1/4 and 1/2 mile radii	-0.065** (0.022)	-0.020* (0.010)	-0.033 (0.019)	-0.012* (0.005)
<i>Low SES Neighborhoods for 65 W&S Stores (N=213,363) (15 Treatment and 50 Control Groups)</i>				
	Total Crime	Violent Crime	Property Crime	Misdemeanor
(1) Within 1/8 mile radius	0.067*** (0.013)	0.013* (0.006)	0.042*** (0.011)	0.010* (0.004)
(2) Within 1/4 mile radius	0.146*** (0.024)	0.015 (0.011)	0.112*** (0.020)	0.020** (0.007)
(3) Within 1/2 mile radius	0.167*** (0.040)	0.019 (0.020)	0.121*** (0.032)	0.039** (0.012)
(4) Between 1/8 and 1/4 mile radii	0.078*** (0.019)	0.001 (0.009)	0.070*** (0.016)	0.008 (0.006)
(5) Between 1/4 and 1/2 mile radii	0.022 (0.030)	0.005 (0.016)	0.006 (0.023)	0.020* (0.009)

Note: Coefficients in cells are average marginal effects of predicted number of events, coming from unique Poisson regressions. See the note of the Table 2-9 for more details. The high/low SES neighborhoods were identified depending on whether an inflation-adjusted median house income of a census 2000 tract in which a W&S store was located was higher than \$50,110 in the 2011 dollars.

Chapter 3 . A BAYESIAN ANALYSIS OF THE PHILADELPHIA PROBATION EXPERIMENT

Abstract

Despite increasing interest in the cognitive behavioral therapy (CBT) treatment-based probation program, relatively little is known about whether this program's effect on recidivism may vary according to probationer characteristics. Using the rich data collected from a multi-year randomized field trial focusing on high risk probationers in Philadelphia and a rigorous Bayesian hierarchical Gamma-Poisson model, the current paper compares probationers who were assigned to the CBT program (treatment group) and those who were not (control group), conditional on probationer characteristics. The results show that the CBT program effect in reducing recidivism is more evident for the high-risk probationers who are at around 10-19 and 30-39 years old, who have ever experienced probation prior to the current probation, and who have a high ratio of "high (risk prediction)" votes from the random forest risk prediction model. Noticeably, the Bayesian approach detects more statistically meaningful intention-to-treat (ITT) effects than the frequentist null hypothesis significance testing (NHST) approach does. This result is due to the Bayesian approach's strength of the relative robustness to small sample sizes and outliers impacts that are common in criminal justice field trials.

3.1. Introduction

Probation is a prominent alternative to incarceration in the U.S. At the end of 2010, the probation population in the U.S. totaled more than 4 million, while the incarceration population was just over 2.2 million (Glaze, 2011). However, there is a long-standing lack of evidence for efficient and effective probation programs (Morgan, 1993; Petersilia, 2011). Historically, evaluations of probation programs have suggested either that “Nothing Works” (Martinson, 1974), or that such programs had high failure rates (General Accounting Office, 1976).

Efforts, however, continue to identify models of probation that will work for individuals at high-risk for recidivism. Recently, cognitive-behavioral therapy programs combined with intensive surveillance (hereafter "CBT-ISP") attract attentions as a promising model for probation (Petersilia, 2011). Meta-analysis evaluations on CBT-ISPs show that the program has positive effects in reducing the recidivism rates of high-risk offenders (Landenberger and Lipsey, 2005; Lipsey and Landenberger, 2005; Lipsey, Chapman, and Landenberger, 2001; Lipsey, Landenberger, and Wilson, 2007).

However, there are relatively few studies of CBT-ISPs in large scale settings that involve experimental designs. In addition, we know little about what characteristics of offenders, if any, are critical to the effective implementation of CBT-ISP. The current paper contributes to the CBT-related body of research by investigating for what types of high-risk offenders a CBT program is effective in reducing recidivism rates. Data are derived from an experiment conducted in Philadelphia. For high-risk offenders who participated in a randomized controlled probation experiment in Philadelphia, this study

found that age, prior probation history, and risk levels were associated with the program effectiveness.

More importantly, this paper contributes to the criminal justice evaluation field by suggesting that a Bayesian hierarchical modeling can lead to statistical conclusions different from those that its corresponding frequentist model makes. When outliers and small samples are present, frequentist models are more likely to falsely fail to reject the null hypothesis because the data heterogeneity may diminish a frequentist approach's statistical power. A Bayesian model approach is potentially preferable to its corresponding frequentist model in that it is relatively less influenced by outliers and small sample sizes. The results in the current paper show that the Bayesian hierarchical Gamma Poisson model detected more meaningful intention-to-treat (ITT) effect differences between the treatment group who were assigned to the CBT-ISP and the control group who were not. This consequence suggests the possibility that some of the past probation or community correction program failures might be not a failure of the program per se but a failure of program effect detection.

The remainder of this paper is organized as follows. Section 2 provides a brief literature review on probation reforms, including description of the literature relevant to a consideration of the factors associated with the success or failure of both probation CBT programs. Section 3 explains the data and variables that are used in the current analysis and the Philadelphia probation experiment. Section 4 details the procedures of the Bayesian hierarchical Gamma Poisson analysis on the frequency data. Results and robustness tests follow in Sections 5 and 6. The paper ends with a discussion and

conclusion in Section 7, mainly focusing on the implication of using a Bayesian modeling in the fields of criminology and criminal justice.

3.2. Literature Review

Research on probation-supervision characteristics that might lead to program failure or success proliferated in 1970s and 1980s, under the belief that probation could be improved if resources were distributed optimally to those who were likely to be responsive to intervention. Previous studies identified that age, gender, race, employment, marital status, prior criminal or arrest records, and the length of supervision were important predictors of probation success (Morgan, 1993).

However, there was no consensus on what characteristics were important and meaningful for program delivery. For example, age was reported to be inversely related with failure in many studies; the younger a probationer was, the more likely the person was to recidivate after probation (Cadwell, 1951; Davis, 1964; Liberton, Silverman, and Blount, 1992; McCarthy and Langworthy, 1987; Wood and O'Donnell, 1980). In particular, Bartell and Winfree Jr. (1977) suggested age 28 as the peak time of probation recidivism; and Cunniff (1987) reported that those in their mid-20s were most likely to fail in probation. In contrast, a number of studies reported no significant relationship between age and probation results (Landis, Mercer and Wolff, 1969; Martin, Cloninger, and Guze, 1978; Roundtree, Edwards, and Parker, 1984; Scott and Covey, 1983). Race was another controversial characteristic. For instance, while Cockerill (1975) reported

race as a critical success characteristic, Roundtree, Edwards, and Parker (1984) found that it played no role in probation success or failure.

Since the 1980s, the intermediate/intensive sanction program (hereafter, "ISP"), which features intensive surveillance and intermediate-level punishments between routine probation sanction and incarceration, has increased in popularity (Petersilia, 1999 and 2011; Petersilia and Turner 1993). Beginning in Georgia in the early 1980s, this model was adopted by every state by 1990, and some of the programs are still valid even today (Petersilia, 2011; Petersilia and Turner, 1993). However, the evaluation results from experiments were not favorable for the ISP intervention. Overall, the ISP rarely decreased the participants' recidivism rates. Rather it resulted in increased technical violations and revocations, thus driving up incarceration rates and the total cost to the correctional systems (Petersilia, 2011; Petersilia and Turner, 1993; for a comprehensive review, see Lipsey and Cullen, 2007, and Smith, Gendreau, and Swartz, 2009).

With increasing criticism of the sole ISP, hybrid community correction models that combine intensive surveillance and well-designed rehabilitative treatment have gained in popularity (Lipsey, 1992; Lösel, 1995). Among the treatments suggested, cognitive-behavioral therapy (CBT), which gives recipients an opportunity to modify abnormal cognitive and behavioral habits through psychology-based classes, is said to be the most promising (Petersilia, 2011). Based on a large number of individual evaluations and meta-analyses to date, the CBT-ISPs have yielded more effective results in reducing recidivism rates than have other treatment programs (Lipsey, 1992; 1995; Lösel, 1995),

by a range of 10% ~27% (Landenberger and Lipsey, 2005; Lipsey and Landenberger, 2005; Lipsey, Chapman, and Landenberger, 2001; Lipsey and Cullen, 2007).

However, despite these promising results, there are still issues that the current CBT-ISP evaluations have not addressed. Despite a rising trend in experiment designs, non-experimental designs are still popular; and the meta-analyses mix experimental and non-experimental results. Rigorous quasi-experimental designs may be informative, but it is not always clear how rigorous the non-experimental designs are in this context. Any meta-analysis can be influenced by the inherent methodological heterogeneity.

Even among the experiments included in the review of CBT-ISP programs, several suffered from small sample sizes, possibly due to the need to constrain resources or maintain a highly controlled experiment. For example, Ross, Fabiano, and Ewles (1988) randomly assigned only 62 offenders to one control and two treatment groups. The small sample sizes in experimental evaluations of CBT-ISP programs means that the results might not be applicable to larger policy context of program delivery to all eligible probationer.

Also, for the CBT-ISPs, relatively few studies investigate which characteristics are closely related with program success or failure. Although the current meta-analyses attempt to identify which characteristics or factors are associated with successful probation by using the meta-regression method, their results are for all probation program that are not confined to the CBT programs (for example, Lowenkamp, Latessa, and Holsinger, 2006). Given the limitations of existing experimental evaluations and the

concern with relatively small samples of subgroups in program evaluations, more knowledge is needed on whether CBT-ISP works in general and for which probationers.

3.3. Data: Philadelphia Probation Experiment

The current paper's data come from the "Philadelphia Anti-Violence Experiment" project, a randomized controlled trial conducted jointly by the Philadelphia Adult Probation and Parole Department (APPD) and the Jerry Lee Center of Criminology (JLC) at the University of Pennsylvania (Hyatt, 2013).⁴⁴ Since the project onset in 2010, all the offenders under the APPD supervision were assigned a risk score. The risk scores, which predicts an offender's likely risk to commit an offense during the two years after assignment to probation, were calculated through the machine-learning-based random forest model that reflected criminal history, prior sentences, and demographic information in forecasting risks (for explanations of the risk classification systems and prediction methods of the random forest model, see Berk et al., 2009). Among the probationers assigned with risk scores, only "high-risk" male offenders, who had "high (risk prediction)" votes over "medium (risk prediction)" and "low (risk prediction)" votes, were enrolled in the experiment.

The Philadelphia experiment had some noticeable strength, compared to other published CBT-ISP experiments. First, the risk assessment tool was based on the random forest model, thus theoretically and practically improving the identification of potential

⁴⁴ I am grateful to Dr. Jordan Hyatt, the Philadelphia Adult Probation and Parole Department, and the Jerry Lee Center of Criminology at the University of Pennsylvania for providing the current subset of Philadelphia probation experiment data for the current study.

"high-risk" offenders. Second, the experiment enrolled a large sample of high-risk probationers making the study more generalizable to program delivery on a large scale. Third, efforts to adhere to the implementation protocols were successful, thus minimizing any possible effect of implementation issues on outcomes (Hyatt, 2013).

The experiment focused on assessing the difference between CBT-ISP and two other forms of probation delivery protocols on a variety of outcomes including: recidivism (total, serious, violent, property, sexual, and drug-related arrests), absconding, and urinal analysis results. Two of the main protocols examined in the experiment were:

1. An intensive supervision (ISP) group, for whom weekly reporting, drug tests, and home visits are required.
2. An intensive supervision with cognitive-behavioral therapy (CBT-ISP) group, for whom monthly reporting, drug testing (if ordered), and attending the CBT classes (14 distinct weekly lessons) are required.

As of April 30th, 2011, 447 high-risk offenders were assigned to the ISP group, and 457 to the CBT-ISP group (Hyatt, 2013).

For the purpose of evaluating the CBT program effectiveness, the current paper compared the ISP (hereafter "control") and CBT-ISP (hereafter "treatment") groups. Comparing the control and the entire treatment groups was associated with estimating the effect of intention-to-treat (ITT) (Gross and Fogg, 2004).⁴⁵⁴⁶

⁴⁵ In addition to the ITT, the effect of the treatment-on-the-treated (ETT), which utilizes only those who actually participate in an intervention program and comply with program protocols, may be estimable.

The outcome in this paper was limited only to numbers of all types of arrest-leading violations that a probationer committed within one year after assignment to a group. Therefore, the current paper aimed to conduct a frequency analysis. To avoid confusion derived from setting up a totally separate statistical model, a prevalence outcome analysis was not included in the current paper. For the given data, Table 3.1 provides summaries of the outcome, numbers of violations committed by one probationer within one year. Overall, the numbers of violations for both the groups were severely skewed to the right, with zero violations of more than 50% probationers per group and a few severe outliers being greater than 100.

Table 3-1. Summary of the Outcome, "Any Violation Committed Within One Year"

	Min.	1 st Quartile	Median	Mean	3 rd Quartile	Max.
1. Control Group (N=447)	0	0	0	4.18	4	117
2. Treatment Group (N=457)	0	0	0	3.70	3	132

Note: Values indicate numbers of violations committed by one probationer within one year.

However, because a control group is not exposed to a treatment, it is not revealed whether a control group member is a potential complier or non-complier. It is known that for a binary outcome case, the ETT estimation is relatively easy because a treatment group's complier proportion can be easily used for inferring the control group's potential complier proportion (Sommer and Zeger, 1991). For a frequency outcome case, however, it seems that there is no consensus yet on an appropriate way to estimate ETT (Efron and Feldman, 1991; Forcina, 2006; Frölich and Melly, 2013; Shpitser and Pearl, 2006). To my knowledge, Imbens and Rubin (1997) and Long, Little, and Lin (2010) are the only studies that estimate the ETT from the Bayesian perspective, using techniques based on a missing data analysis and a data augmentation algorithm. However, the current paper confines its main window to the ITT effect estimates only, because 1) the Bayesian ETT estimation methods of Imbens and Rubin (1997) and Long et al. (2010) require different models and assumptions than those of the current study, and 2) ITT estimates can demonstrate the contrast between the Bayesian and frequentist approaches.

⁴⁶ Appendix 3-A provides a simple simulation result for a treatment-on-the-treated effect (ETT) estimation using the current Bayesian Monte Carlo grid model. However, based on strong assumptions, the results should be limitedly interpreted as just a preliminary and exploratory one.

The control and treatment groups were broken into subgroups according to four supervision characteristics: age, race, previous probation experience, and a ratio of the random forest high-risk votes. The control and treatment subgroups in the same characteristic category were compared then.⁴⁷ "Age" indicated the age at which a probationer was randomly assigned to a group. These ages were categorized into five ten-year age subgroups, from "age 10-19" to "age 50 or older." "Race" had the three categories of white, black, and Asian. Hispanic ethnicity was ignored due to too low a volume (N=4). A previous probation experience literally referred to whether a high-risk offender experienced any probation before the current probation program. According to such an experience, probationers belonged to either of the two subgroups: the "not experienced" and "experienced."

A ratio of random forest "high" votes is the ratio of "high" votes over the total 500 high/medium/low votes that were obtained from the Random forest risk-forecasting procedure (Berk et al., 2009). For the comparison purpose, based on whether the ratio was equal to or greater than the total mean ratio ($=0.4255$), probationers were assigned to either of two subgroups: the "low ratio of high votes," whose ratios were less than 0.4255, and the "high ratio of high votes," whose ratios were equal to or greater than 0.4255.

Table 3.2 reports the proportions of the four supervision characteristic categories between the control and treatment groups. Overall, due to the random assignment, each of the characteristic categories was well-balanced between the two groups, occupying

⁴⁷ Notice that it did not mean that subgroups were compared over characteristic categories. Although the control and treatment subgroups were comparable for a given characteristic category because they were well-balanced due to the random assignment, the causal effect of different characteristic categories could not be estimated across the control or treatment groups because they were not randomly assigned. .

similar proportions of the total. For example, the proportions of those whose ages at assignment were 20-29 are 48.1% for the control group and 47.7% for the treatment group.

Table 3-2. Proportions of Probationer Characteristic Categories

	Control Group (ISP group)	Treatment Group (CBT-ISP group)
<u>1. Age</u>	<u>100.0%</u> (N=447)	<u>100.0%</u> (N=457)
1.-(1) % Age 10-19	13.7% (N=61)	10.7% (N=49)
1.-(2)% Age 20-29	48.1% (N=215)	47.7% (N=218)
1.-(3) % Age 30-39	21.0% (N=94)	22.1% (N=101)
1.-(4) % Age 40-49	13.4% (N=60)	14.0% (N=64)
1.-(5) % Age 50 and older	3.8% (N=17)	5.5% (N=25)
<u>2. Race</u>	<u>100.0%</u> (N=447)	<u>100.0%</u> (N=457)
2.-(1) % White	21.0% (N=94)	22.5% (N=103)
2.-(2) % Black	71.8% (N=321)	70.5% (N=322)
2.-(3) % Asian	4.3% (N=19)	4.6% (N=21)
2.-(4) % (Missing Values and Others)	2.9% (N=10)	2.4% (N=11)
<u>3. Prior Probation Experience</u>	<u>100.0%</u> (N=447)	<u>100.0%</u> (N=457)
3.-(1) % Not Experienced	37.4% (N=167)	32.0% (N=146)
3.-(2) % Experienced	62.6% (N=280)	68.1% (N=311)
<u>4. Ratio of Random Forest "High" Votes</u>	<u>100.0%</u> (N=447)	<u>100.0%</u> (N=457)
4.-(1) % Low Ratio of "High" Votes	52.6% (N=235)	54.3% (N=248)
4.-(2) % High Ratio of "High" Votes	47.4% (N=212)	45.7% (N=209)

Note: Sample sizes are reported in parentheses.

3.4. Statistical Modeling

3.4.1. Research Questions

The current paper focuses mainly on the following three questions:

1. From a Bayesian perspective, was the Philadelphia CBT-ISP probation protocol successful in reducing recidivism rates, compared to the ISP only protocol?
2. When the entire population was broken into subgroups according to characteristics, for which subgroups was the Philadelphia CBT-ISP probation protocol successful in reducing recidivism rates, compared to the ISP only protocol, from a Bayesian perspective?
3. Do the Bayesian evaluation results differ from those of the frequentist evaluation?

3.4.2. Bayesian Analysis in Criminology

The Bayesian approach is based on the Bayes theorem, $P(\theta|Y) = \frac{P(Y|\theta) \cdot P(\theta)}{P(Y)}$, where

θ and Y are two different parameters, and $P(Y) = \int_{-\infty}^{\infty} P(Y|\theta) \cdot P(\theta) d\theta$. From the theorem, the posterior $P(\theta|Y)$ can be rewritten as being proportional to the likelihood $P(Y|\theta)$ multiplied by the prior $P(\theta)$:

$$P(\theta|Y) \propto P(Y|\theta) \cdot P(\theta)$$

The Bayesian approach differs considerably from the frequentist, Null Hypothesis Significance Testing (NHST) approach. Not all the differences will be introduced in this

paper (see Kruschke, 2011a and 2011b; Gelman et al., 2004 for the differences between the Bayesian and the NHST; also see Cohen, 1994; Howard, Maxwell, and Fleming, 2000 for the drawbacks of the NHST), but two fundamental differences are worth noting.

Above all, the Bayesian approach recognizes the uncertainty within a parameter. Unlike the frequentist approach, in which a parameter has a fixed but unknown value, the Bayesian approach allows a parameter to have its own distribution. Bayesian statistical conclusions do not rely on the p-value, which actually indicates the chance probability to observe the current results under the null hypothesis that the parameter has a null value and is commonly used in the NHST approach as the decision criterion. Instead, the Bayesian approach attempts to estimate directly the distribution of a parameter, which is often called "credible interval," on which a statistical conclusion is made. One advantage of this feature is that a small sample size is no longer problematic, as it would be in the frequentist NHST approach. The small sample size only leads to a wide Bayesian credible interval.⁴⁸

Another notable difference is that the Bayesian approach has a prior, $P(\theta)$. Using a prior implies that possibly important information that is obtained a priori influences a statistical conclusion. With more affluent information than found in frequentist models, Bayesian models can improve the quality of statistical conclusions. However, the role of a prior practically depends on the relative role of the corresponding likelihood, $P(Y|\theta)$. If a data size is sufficiently large, the likelihood will dominate the posterior, and the prior

⁴⁸ Because a Bayesian analysis can also be biased if an informative prior distribution is used when a sample size is small, the current study excludes any informative prior.

will rarely influence the posterior. On the other hand, if the data size is small, the prior will play an important role in the posterior construction.

While it has been more than 20 years since the Bayesian approach was introduced in the field of criminology (Berk et al., 1992a and 1992b), only a few studies have followed the lead so far. The two papers from Berk and his colleagues estimate effects of the Colorado Springs Spouse Abuse Experiment on domestic violence reduction through a Bayesian logit regression modeling, while using results of the three previous similar field experiments from Omaha, Milwaukee, and Dade County in Florida as priors. Cohen et al. (1998) use the Bayesian hierarchical Gamma Poisson model, thus setting the stage for the current paper, to see whether arrest rates for drug offenders differed from those of non-drug felony offenders with 1986 and 1990 Los Angeles conviction data.⁴⁹

However, the Bayesian approach has recently attracted greater attention. Sullivan and Mieczkowski (2008) report empirical Bayesian analysis results on whether the ISP were effective in reducing recidivism rates by accumulating extant results of four different location studies that were extracted from Petersilia and Turner (1993) as their priors. Anwar and Loughran (2011) employs a Bayesian updating modeling to develop a learning model which tests how individuals update risk perceptions over time when they receive risk signals in offending, with court data from Maricopa County, AZ, and Philadelphia County, PA. Deller, Amiel, and Deller (2011) use a Bayesian model-

⁴⁹ In serving as the model for the current paper, Cohen et al. (1998) utilized a similar hierarchical Gamma Poisson model. However, while the current paper uses the Monte Carlo grid sampling method, they used the "Metropolis within Gibbs" method to get λ values. Also, while the current paper uses randomized controlled experiment data, they used random samples of administrative data. Because different types of arrest may reflect different inherent characteristics in the samples, despite random sampling, their simple comparison results were relatively weak in addressing any causal relationship.

averaging method to identify which factors are critical in determining rural crime rates in the Midwestern areas. In addition, Law, Quick, and Chan (2014) apply a Bayesian spatiotemporal modeling for a crime-pattern analysis. They use a hierarchical logit regression model with property crime data from the York region of Ontario, Canada.

3.4.3. Hierarchical Poisson Model

This paper basically follows the Cohen et al. (1998)'s the Bayesian hierarchical Gamma Poisson modeling but uses a different Monte Carlo estimation method. Let Y_i be a number of discrete violations that were committed independently of the time since last violation within a fixed time of one year. Y_i of an individual probationer i ($i=1, 2, \dots, N$) is assumed to be Poisson-distributed with a nonnegative parameter λ . This λ value indicates the average rate of the violations, which is of main interest in this paper. However, given the possibly different offending propensity of an individual probationer, λ is now allowed to vary at a level of individual i , being λ_i , where $i = 1, 2, 3, \dots, N$. Also, to enjoy the convenience of the conjugacy property of a Poisson distribution, let λ_i follow a gamma distribution with nonnegative parameters α and β , where α is a shape and β is a scale. The following is the hierarchical Poisson model that has the three unknown parameters, λ_i , α , and β .

$$(1) \quad Y_i | \lambda_i \sim \text{Poisson}(\lambda_i), \text{ where } \lambda_i > 0 \text{ and } \lambda_i \text{ is the expected mean of } Y_i$$

$$(2) \quad \lambda_i | \alpha, \beta \sim \text{Gamma}(\alpha, \beta), \text{ where } \alpha > 0 \text{ and } \beta > 0$$

From (1) and (2), respectively,

$$(3) \quad P(Y_i | \lambda_i) \propto \{ \lambda_i^{Y_i} \cdot e^{-\lambda_i} \}$$

$$(4) \quad P(\lambda_i | \alpha, \beta) \propto \left\{ \frac{\beta^{-\alpha} \cdot \lambda_i^{\alpha-1} \cdot e^{-\lambda_i/\beta}}{\Gamma(\alpha)} \right\},$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \cdot e^{-t} dt$ is a gamma function

3.4.4. Prior

Given that $P(\lambda_i, \alpha, \beta) = P(\lambda_i | \alpha, \beta) \cdot P(\alpha, \beta)$, a prior for the two unknown parameters, $P(\alpha, \beta)$, needs to be set. A prior is a belief distribution on interested parameters that can be obtained without referring to a current data set, denoted desirably in terms of probability densities.⁵⁰ A prior can reflect any belief that does not incorporate current data information, such as previous research results, qualitative statement, and even the analyst's subjective judgment as long as it is reasonably justified.

In this paper, however, a noninformative prior is preferred to a subjective one.⁵¹ A noninformative prior is often described as a "reference" (Bernardo, 1979) or "objective" (Berger, 2006) prior (Gelman et al., 2004). This type of prior is a function that maximizes some measure of distance between a prior and a posterior such that the prior plays a minimal role, but a data-likelihood dominates in forming the posterior belief distribution.

⁵⁰ The distinction between the prior and posterior terms relies solely on whether particular data information of current interest is excluded (prior) or included (posterior), without referring to the time ordering of belief formation (Kruschke, 2011b).

⁵¹ An alternative can be to use the recidivism rates of CBT program probationers in other places. However, comparing recidivism rates between different places may be problematic because each place has a different correctional system in criteria, law-enforcement level, and probation officer discretion level (Petersilia, 1997).

This paper employs the following noninformative prior that Cohen et al. (1998) suggested for a gamma-Poisson hierarchical model. Given little information about behaviors of gamma parameters α and β , they posed a uniform or flat density (which is proportional to 1) on the shrinkage weight Δ in $E(\lambda_i|Y_i)=\Delta \cdot Y_i+(1-\Delta) \cdot \mu$, where $\Delta=\frac{\beta}{\beta+1}$, and μ is a prior mean, $\alpha\beta$, which is >0 . They also placed a flat density on $\delta=\log(\mu)$, where $\mu=\alpha\beta >0$. Therefore, given $\alpha=e^\delta$ and $\beta=\frac{\Delta}{1-\Delta}$, $f(\delta, \Delta) \propto 1$ can be rewritten as follows:

$$(5) \quad P(\alpha, \beta) \propto \frac{1}{\alpha(1+\beta)^2}$$

In addition to Cohen et al.'s prior, a Jeffreys prior and a flat prior on $P(\alpha, \beta)$ are also considered as noninformative priors in the robustness check section. A Jeffreys prior, suggested first by Jeffreys (1961), is a prior that the logarithms of parameters are expected to have uniform densities when parameters are independent and range in $(0, \infty)$, thus yielding results that are invariant to a scale change due to parameter transformations. The Jeffreys (6), which is excerpted from Miller (1980), and flat (7) priors for a gamma distribution are given as follows:

$$(6) \quad P(\alpha, \beta) \propto \frac{\{\alpha\psi'(\alpha)-1\}^{1/2}}{\beta}, \text{ where } \psi' = \frac{\partial^2}{\partial \alpha^2} \log(\Gamma(\alpha))$$

$$(7) \quad P(\alpha, \beta) \propto 1$$

3.4.5. Posterior via the Monte Carlo grid sampling Method

From (3) and (4), a posterior $P(\underline{\lambda}, \alpha, \beta | \underline{Y})$, where $\underline{\lambda} = \lambda_1, \lambda_2, \dots, \lambda_N$ and $\underline{Y} = Y_1, Y_2, \dots, Y_N$, can be described as follows:

$$\begin{aligned}
P(\underline{\lambda}, \alpha, \beta | \underline{Y}) &\propto P(\underline{Y} | \underline{\lambda}, \alpha, \beta) \cdot P(\underline{\lambda}, \alpha, \beta) \\
&\propto P(\underline{Y} | \underline{\lambda}) \cdot P(\underline{\lambda} | \alpha, \beta) \cdot P(\alpha, \beta) \\
&\propto \prod_{i=1}^N P(Y_i | \lambda_i) \cdot \prod_{i=1}^N P(\lambda_i | \alpha, \beta) \cdot P(\alpha, \beta)
\end{aligned}$$

By (3), (4), and (5),

$$(8) \quad P(\underline{\lambda}, \alpha, \beta | \underline{Y}) \propto \frac{1}{\alpha \cdot (1+\beta)^2} \cdot \prod_{i=1}^N \left\{ \frac{\beta^{-\alpha}}{\Gamma(\alpha)} \cdot \lambda_i^{Y_i + \alpha - 1} \cdot e^{-\left(\frac{\beta}{1+\beta}\right)\lambda_i} \right\}$$

This (8) can be integrated over λ for $P(\alpha, \beta | Y)$:

$$\begin{aligned}
P(\alpha, \beta | \underline{Y}) &= \int_0^\infty P(\lambda, \alpha, \beta | Y) d\lambda \\
&\propto \frac{1}{\alpha(1+\beta)^2} \cdot \prod_{i=1}^N \left\{ \left(\frac{\beta^{-\alpha}}{\Gamma(\alpha)} \right) \cdot \int_0^\infty (\lambda_i^{Y_i + \alpha - 1} e^{-\left(\frac{\beta}{1+\beta}\right)\lambda_i}) d\lambda_i \right\} \\
(9) \quad P(\alpha, \beta | \underline{Y}) &\propto \frac{1}{\alpha(1+\beta)^2} \cdot \prod_{i=1}^N \left\{ \left(\frac{\beta^{-\alpha}}{\Gamma(\alpha)} \right) \cdot \left(\frac{\beta}{1+\beta} \right)^{Y_i + \alpha} \cdot \Gamma(Y_i + \alpha) \right\}
\end{aligned}$$

Because this joint distribution of the two unknown parameters α and β in (9) is not a standard distribution form, the respective conditional distributions of α and β cannot be derived analytically.

However, a Monte Carlo method, where an unknown parameter's distribution is estimated by random samples from numerous repeated simulations, can yield acceptably accurate conditional distributions of α and β , respectively. Given that there are only two unknown parameters of α and β , an efficient Monte Carlo grid method is employed.⁵² The Monte Carlo grid procedure for the two Gamma α and β parameters is

⁵² As a general form of a grid method, the Gibbs Sampler (Geman and Geman, 1984) or the Metropolis algorithm (Metropolis et al., 1953) may be used instead for the current purpose of estimating α and β distributions (see Cohen et al., 1998, for an example). However, a grid method can be more efficient than the general approaches for two parameter estimation as long as a grid net is appropriately selected (Kruschke, 2011b; Gelman et al. 2004).

described in the following four steps (Ritter and Tanner, 1992:863, Gelman et al., 2004:91-92);

- (A) Appropriately choose the grid net of (α, β) points, where $0 < \alpha_{\min} \leq \alpha \leq \alpha_{\max}$ and $0 < \beta_{\min} \leq \beta \leq \beta_{\max}$. On the chosen rectangular grid net of (α, β) points, evaluate (9) to obtain the unnormalized discrete joint densities of (α, β) . This evaluation constructs contour lines of equal densities.
- (B) Normalize the discrete joint densities; divide them by the total sum of densities so that the total probability amounts to 1. Then, compute the discrete conditional distribution of α , $P(\alpha | Y)$ by summing over β on the grid. Given an α value on the grid, also compute the discrete conditional distribution of β , $P(\beta | \alpha, Y)$.
- (C) With a discrete probability of $P(\alpha = \alpha^1 | Y)$, randomly choose one α value, α^1 between α_{\min} and α_{\max} . Also, with a discrete probability of $P(\beta = \beta^1 | \alpha, Y)$, randomly choose one β value, β^1 between β_{\min} and β_{\max} . Let the chosen value pair be (α^1, β^1) .
- (D) Repeat step (C) by K times so that K pairs of (α^J, β^J) , where $J=1, 2, \dots, K$, are obtained. The K pairs of (α^J, β^J) that are obtained by the MCMC method can well represent the continuous joint probability of $P(\alpha, \beta | Y)$ in (9).

Given the K pairs of (α^J, β^J) from step (D), $\lambda_i | \alpha, \beta$ can be obtained:

$$(10) \quad \lambda_i^J | \alpha^J, \beta^J \sim \text{Gamma}(\alpha^J, \beta^J), \text{ where } J=1, 2, \dots, K$$

Randomly choose a $\lambda_i^J | \alpha^J, \beta^J$ value from a gamma distribution of $\text{Gamma}(\alpha^J, \beta^J)$ in

(10). Repeat this random choice of λ_i^J for K times, so that K values of λ_i^J are obtained.

These K λ_i^J values obtained via the MCMC method can well represent the conditional distribution of λ_i , $P(\lambda_i | \alpha, \beta)$.

3.4.6. Bayesian Approach for a Program Evaluation

Given $i=1, 2, \dots, N$ and $J=1, 2, 3, \dots, K$, total $N \times K$ values of λ_i^J have been estimated so far. To obtain the group parameter λ that is a group expected mean of Y,

$$(11) \quad \lambda^J = \frac{1}{N} \sum_{i=1}^N \lambda_i^J \quad (J=1, 2, \dots, K)$$

This simple average indicates that each individual i 's λ_i^J is weighted equally. The K values of λ^J represent the distribution of the group parameter λ . The mean of the K values of λ^J is described as the mean of the group parameter λ , and the 95% Highest Density Interval (HDI)⁵³ of the K values of λ^J is regarded as the 95% HDI of λ .

Given the fact that the current data are obtained from the large-scale randomized controlled experiment, a simple mean parameter comparison between the

⁵³ A 95% Highest Density Interval (HDI) indicates a 95% range of an entire parameter distribution such that parameter values within the 95% range have higher believability to be realized than those within the remaining 5% range (Kruschke, 2011b). A Bayesian interval such as the HDI is strong enough to provide distributional information; with uni-modality, the closer to the center of the interval a value is located, the higher believability the value has. Therefore, the central tendency measure can bear the highest believability among possible values (*ibid.*). Note also that because a 95% HDI is a collection of realizable parameter values, the 95% HDI actually indicates that the probability that the interval include a true value is 95%, unlike the frequentist approach's confidence interval.

control and treatment groups can be a reasonable evaluation tool. By random assignment, all known and unknown characteristics between the two groups are expected to be balanced.

The group-level λ contrasts (differences) between the treatment and control groups, $(\lambda_{Treatment} - \lambda_{Control})$, can be obtained from K values of $(\lambda_{Treatment}^J - \lambda_{control}^J)$, where $(\lambda_{Treatment}^J - \lambda_{control}^J) = \frac{1}{N} \sum_{i=1}^N (\lambda_{i,Treatment}^J - \lambda_{i,control}^J)$ (J=1,2,...,K). From these values, the mean and 95% HDI of the K values of $(\lambda_{Treatment} - \lambda_{Control})$ can be calculated. It is these mean and 95% HDI of the λ contrasts that provide evaluation judgment information about whether and how much the treatment group meaningfully differs from the control group. When the 95% HDI includes zero, the treatment group will be judged to be the same as the control group. Meanwhile, if the 95% HDI excludes zero, the treatment group will be judged to differ from the control one. The mean will be regarded as the highest believable amount of the difference between the groups.

The number N indicates a sample size of a subgroup, thus varying according to which subgroup is being considered. The number K, which means the iteration number of simulations, is fixed to 1,000 in this paper. All the current analysis is conducted through the R program (R version 3.0.2.).

3.4.7. Alternative: Frequentist Approach with Null Hypothesis Significance

Test (NHST)

Instead of the suggested Bayesian method, a frequentist approach is also available for the CBT program evaluations. As noted in 4.1., one of the major differences between the Bayesian and the frequentist approaches comes from the assumption of a parameter. While Bayesians think that a parameter has its own distribution with uncertainty, frequentists regard a parameter value as being fixed but unknown. This difference results in different models. The Poisson parameter λ_i , where i is an individual probationer = 1, 2, ..., N , cannot take its own additional distribution in a frequentist modeling. Instead, given that there is only one Y_i for an individual i in the current setting, the parameter λ_i for an individual i can be estimated through the maximum likelihood method as follows:

$$\begin{aligned} L(\lambda_i) &= Y_i \cdot \log(\lambda_i) - \lambda_i - \log(Y_i!), \text{ where } L(\lambda_i) \text{ means Log-Likelihood}(\lambda_i) \\ L'(\lambda_i) &= \frac{Y_i}{\lambda_i} - 1 = 0 \\ (11) \quad \widehat{\lambda_{i,MLE}} &= Y_i \end{aligned}$$

In other words, each individual's data Y_i is the best estimation of each individual's violation tendency, λ_i . If the group mean λ is an average of all λ_i values with equal weights,

$$(12) \quad \widehat{\lambda_{MLE}} = \sum_{i=1}^N \widehat{\lambda_{i,MLE}} = \sum_{i=1}^N Y_i = \bar{Y}$$

The point estimation of λ is the group sample mean of Y_i 's, \bar{Y} . Also, its corresponding 95% confidence interval estimation of λ can be obtained as follows, with an assumption that Y_i 's sampling distribution is the t-distribution:

$$(13) \quad \bar{Y} \pm t_{0.975, N-1} \cdot \frac{s}{\sqrt{N}}, \text{ where } s \text{ is the sample standard deviation of } Y_i \text{'s.}$$

Under the same logic, the point and 95% confidence interval estimation for $(\widehat{\lambda_{MLE, Treatment}} - \widehat{\lambda_{MLE, Control}}) = (\overline{Y_{Treatment}} - \overline{Y_{Control}})$ can be obtained. In the frequentist approach, evaluation judgments can be made being based on t-test p-values.⁵⁴ When a p-value for a difference between the control and treatment groups is less than 0.05, the null hypothesis that there is no difference between the treatment and control groups is rejected, thus concluding that the treatment was effective in reducing recidivism rates. Being a p-value < 0.05 exactly corresponds to the situation where a 95% confidence interval includes zero. Meanwhile, when a p-value is equal to or greater than 0.05 or when a 95% confidence interval excludes zero, the null hypothesis cannot be rejected, thus leading to the conclusion that the treatment had no effect. It is worth noting that the 95% confidence interval can be hugely influenced by s and N values.

3.5. Results

Table 3.3 reports the Bayesian means and 95% HDIs of λ (in part I) and λ contrasts between the control and treatment groups (in part II). The 95% HDI indicates the credible interval of the most likely 95% λ values that are obtained from the Monte Carlo grid procedure, and the mean indicates the average of the 95% HDI λ values. For

⁵⁴ Alternatively, an evaluation judgment can be made based on a simple Poisson regression's coefficient result, where Y is the outcome and a dichotomous variable indicating the treatment or control group is the regressor. The judgment logic remains the same as that of the t-test. However, the Poisson regression coefficient corresponds to the multiplicative factor between the treatment and control group outcomes, $\frac{Y_T}{Y_C}$, while the t-test coefficient corresponds to the additive difference between the two groups, $Y_T - Y_C$.

clear comparisons, results from the frequentist (NHST) approach are also annotated next to the Bayesian results.

Section I in Table 3.3 reports the levels of estimated group means of violations ($=\hat{\lambda}$). For the control group, who were intensively supervised but not provided with the CBT lessons, the group mean was estimated to be 4.51, while its 95% HDI ranged between 4.30 and 4.73. Compared to the control group, the treatment group had lower levels of estimated group means of violations of 3.95 with the 95% HDI range from 3.78 to 4.15. As can be easily expected, the treatment had a noticeably lower level of the mean than that of the control group.

Section II in Table 3.3 shows the λ contrasts (differences) between the control and treatment groups, which are of the main interest in this paper. The entire 95% HDI of the λ contrasts between the control and treatment groups (-0.83, -0.25) fell below zero. The interval had the estimated group mean of -0.56, meaning that a probationer in the treatment group was on average by 0.56 units less likely to commit a violation than one in the control group at the 95% credible level. Therefore, the CBT program was effective in reducing recidivating behaviors of high-risk offenders in terms of the effect of intention-to-treat (ITT).

It is worth noting again that these λ values were obtained from the Monte Carlo procedure, based on the two Gamma α and β parameters through the grid method. Figure 3.1 depicts the two-dimensional grid that was used in the Monte Carlo grid method for the λ estimation of the control group (N=447) (see the Monte Carlo grid procedure [A] in Section 3.4.5). The α grid was set to range from $\alpha_{\min}=0.13$ to $\alpha_{\max}=0.20$, while the β grid

was set from $\beta_{\min}=19$ to $\beta_{\max}=37$. The contour lines indicate the equal values of unnormalized discrete joint densities of (α, β) . Based on the contours in Figure 3.1, the conditional probability $P(\alpha | Y)$ and $P(\beta | \alpha, Y)$ were calculated (see the Monte Carlo grid procedure [B] in Section 3.4.5.), as depicted in Figures 3.2 and 3.3, respectively.⁵⁵

On the other hand, the results from the frequentist NHST perspective showed somewhat different evaluation results than those from the Bayesian perspective. The point estimations of λ_{MLE} appeared to be very similar to those of the Bayesian λ group means. However, the 95% confidence intervals were wider in the absolute width than the Bayesian 95% HDIs, although the two intervals have different meanings.⁵⁶ For the contrast between the treatment and control groups, the 95% frequentist confidence interval, $(-1.94, 0.99)$, was wider than the Bayesian HDI and included a zero value in it. Therefore, according to the frequentist approach, the evaluation conclusion would be that the CBT program was not effective in reducing recidivism of high-risk offenders when the effect was measured in the ITT term.

⁵⁵ Every Bayesian estimation of λ in this paper has a separate set of its contour plot like Figure 3.1, and $P(\alpha | Y)$ and $P(\beta | \alpha, Y)$ plots like Figures 3.2 and 3.3. They are not reported, however, in this paper due to space limitations. The collections of contour plots, $P(\alpha | Y)$, and $P(\beta | \alpha, Y)$, for all the λ estimations (total 13 characteristics \times 3 plots \times 3 priors = 117 plots) are available upon request.

⁵⁶ As is well known, a 95% confidence interval, unlike a 95% HDI, cannot be interpreted as the probability that the interval includes a true parameter value within it is 95%.

Table 3-3. Mean and Intervals of λ Levels and λ Contrast for All Treatment and Control Groups

	Bayesian Approach		Frequentist Approach	
	Mean	95% HDI of λ (Lower Bound, Upper Bound)	Point Estimation of λ_{MLE}	95% CI of λ_{MLE} (Lower Bound, Upper Bound)
I. Estimates of λ Levels				
1 Control Group (N=447)	4.51	(4.30, 4.73)	4.18	(3.05, 5.31)
2. Treatment Group (N=457)	3.95	(3.78, 4.15)	3.70	(2.77, 4.64)
II. Estimates of Contrast, $\lambda_{Treatment} - \lambda_{control}$				
2. - 1. Contrast	-0.56	(-0.83, -0.25)*	-0.48	(-1.94, 0.99)

Note: * Indicates, from the Bayesian perspective, that the most likely 95% λ contrast estimates between the control and treatment groups do not include zero, and therefore the two groups differ.

indicates, from the frequentist (NHST) perspective, that the null hypothesis that the difference between the control and treatment groups equals zero is rejected at the 95% confidence level.

Figure 3-1. Contours of Unnormalized Joint Densities of (α, β) for the Control Group

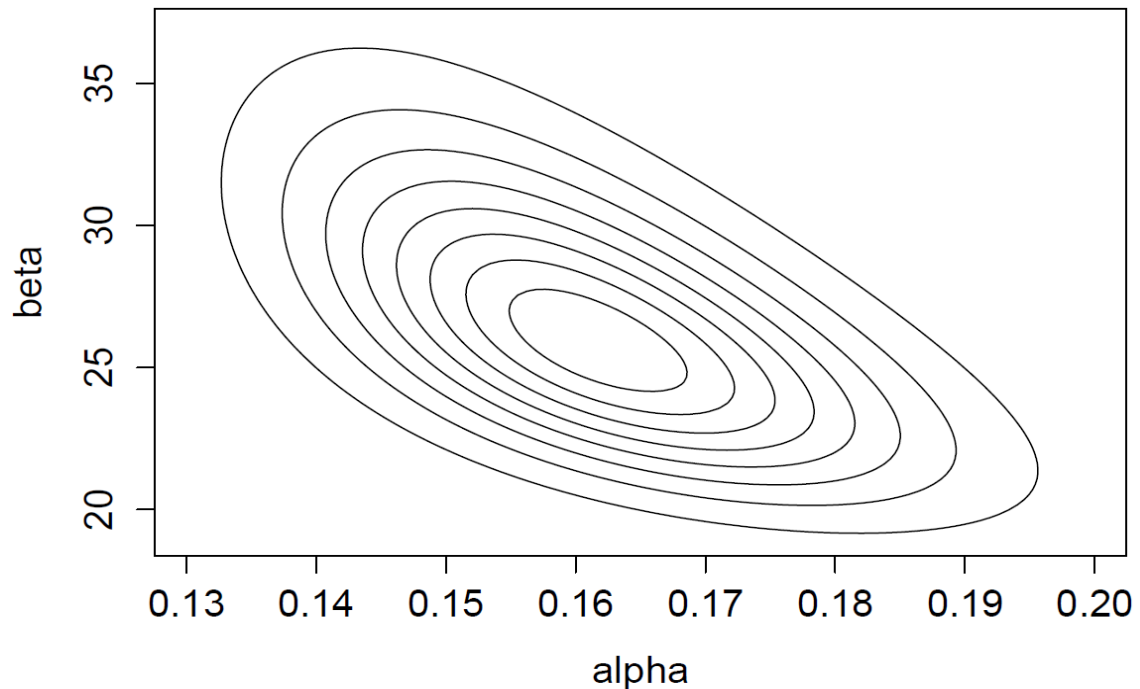


Figure 3-2. Conditional Distribution of α | Y for the Control Group (N=447)

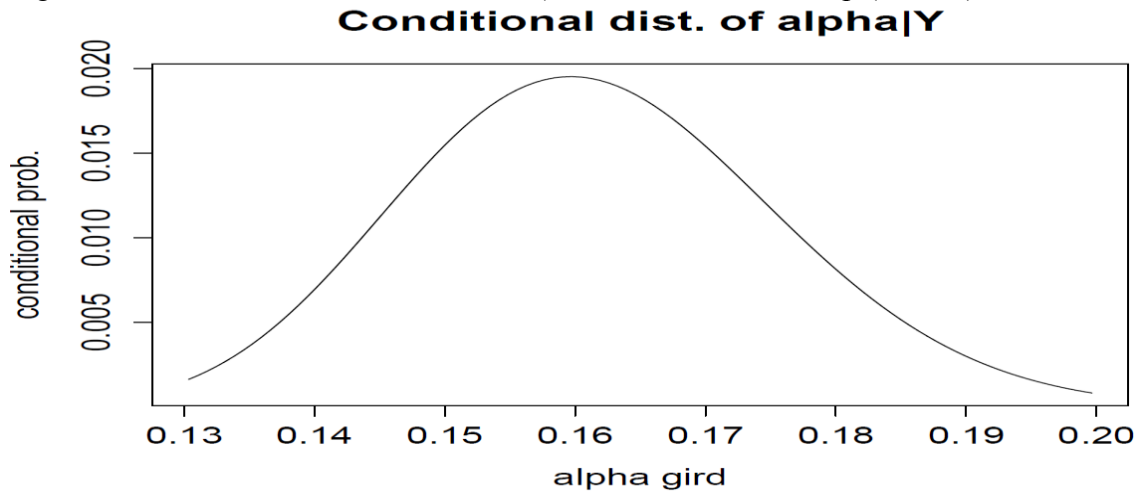


Figure 3-3. Conditional Distribution of β | Y, α for the Control Group (N=447)

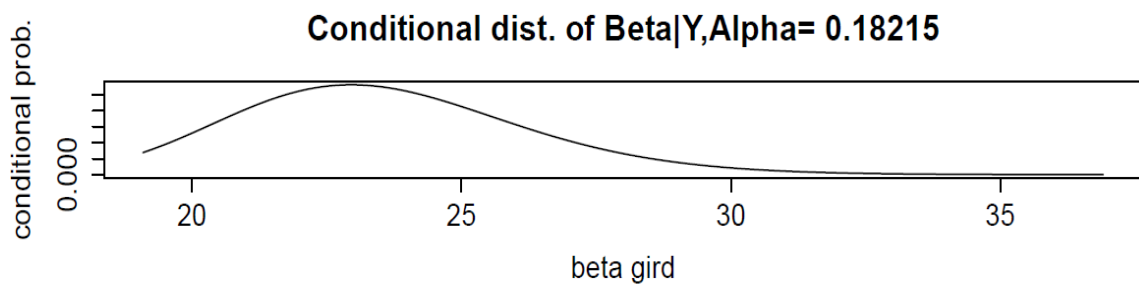
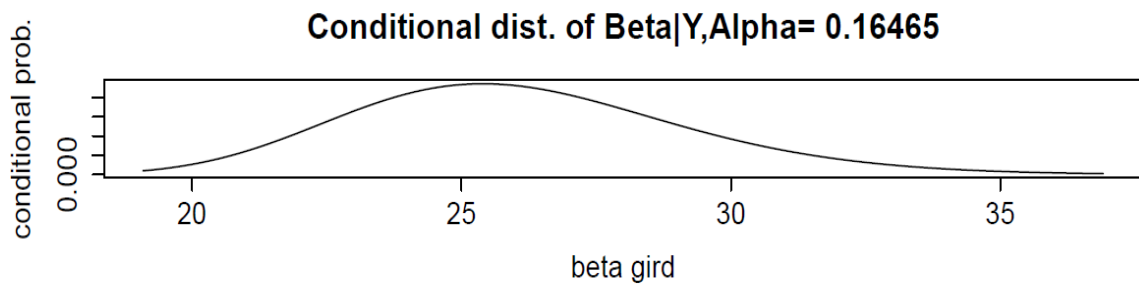
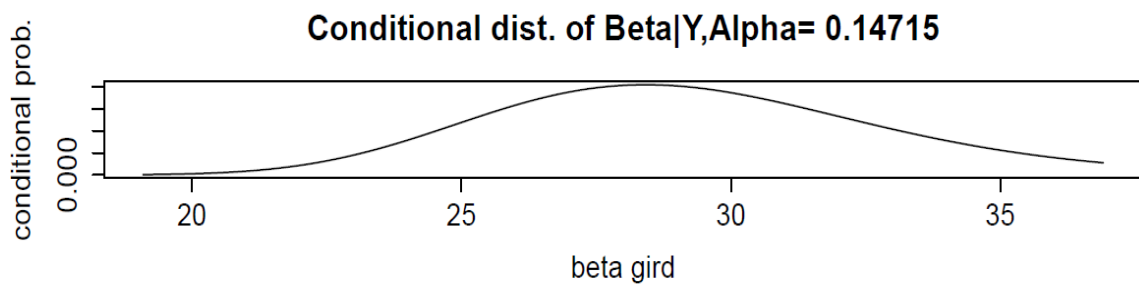


Table 3.4 has basically the same structure as that of Table 3.3. However, this table reports the results when the entire samples were broken down into subgroups according to ten-year age intervals. Notice again that the control and treatment groups were compared, not different age groups were compared to each other. According to part I, the levels of λ estimates were, in general, large for young probationer groups and small for old probationer groups, except for the group age 50 or older. These patterns mesh with the extant research studies that criminogenity peaks in ages of 20s and then decreases (Bartell and Winfree, 1977; Cuniff, 1987; Farrington, 1988).

The estimates of the λ contrast between the control and treatment groups in part II reported somewhat different patterns between the Bayesian and frequentist models. When the Bayesian model was applied, it was evident that the CBT program was effective in reducing recidivating behaviors for the probationer groups ages 10-19 and 30-39 in terms of the ITT effect. For those of the other ages, the CBT program appeared not to be effective in reducing the recidivating behaviors even from the Bayesian perspective.

However, the results from the frequentist t-test model showed that the CBT program was not effective in reducing recidivating behaviors at all for all the age groups. Therefore, according to the frequentist perspective, it is likely concluded that the CBT program had no effect.

Table 3-4. Mean and Intervals of λ Levels and λ Contrast, According to Age Subgroups

	Bayesian Approach		Frequentist Approach	
	Mean	95% HDI of λ (Lower Bound, Upper Bound)	Point Estimation of λ_{MLE}	95% CI of λ_{MLE} (Lower Bound, Upper Bound)
I. Estimates of λ Levels				
I-(1) Ages 10-19				
1. Control Group (N=61)	6.96	(6.20, 7.66)	6.38	(3.49, 9.26)
2. Treatment Group (N=49)	4.11	(3.48, 4.66)	3.88	(1.14, 6.61)
I-(2) Ages 20-29				
1. Control Group (N=215)	4.89	(4.59, 5.20)	4.61	(2.74, 6.48)
2. Treatment Group (N=218)	5.23	(4.90, 5.55)	4.94	(3.20, 6.67)
I-(3) Ages 30-39				
1. Control Group (N=94)	3.85	(3.42, 4.31)	3.47	(0.89, 6.05)
2. Treatment Group (N=101)	2.47	(2.18, 2.83)	2.25	(1.13, 3.36)
I-(4) Ages 40-49				
1. Control Group (N=60)	2.52	(2.05, 3.11)	2.07	(1.06, 3.07)
2. Treatment Group (N=64)	2.47	(1.99, 2.92)	2.06	(0.89, 3.24)
I-(5) Ages 50 and older				
1. Control Group (N=17)	3.07	(1.86, 4.18)	2.29	(0.46, 4.13)
2. Treatment Group (N=25)	3.05	(2.30, 3.83)	2.68	(-0.11, 5.47)
II. Estimates of Contrast, $\lambda_{Treatment} - \lambda_{control}$				
II-(1) Ages 10-19				
2. - 1. Contrast	-2.85	(-3.81, -1.89)*	-2.50	(-6.43, 1.43)
II-(2) Ages 20-39				
2. - 1. Contrast	0.34	(-0.13, 0.79)	0.33	(-2.21, 2.87)
II-(3) Ages 30-39				
2. - 1. Contrast	-1.38	(-1.95, -0.89)*	-1.22	(-4.02, 1.58)
II-(4) Ages 40-49				
2. - 1. Contrast	-0.06	(-0.77, 0.65)	-0.004	(-1.54, 1.53)
II-(5) Ages 50 and older				
2. - 1. Contrast	-0.02	(-1.50, 1.37)	0.39	(-2.87, 3.64)

Note: * Indicates, from the Bayesian perspective, that the most likely 95% λ contrast estimates between the control and treatment groups do not include zero, and therefore the two groups differ.

indicates, from the frequentist (NHST) perspective, that the null hypothesis that the difference between the control and treatment groups equals zero is rejected at the 95% confidence level.

Table 3.5 reports the effectiveness of the CBT program for the race subgroups. Part I in the table indicates that the mean levels of violations were relatively higher for black probationers than for white ones, while the treatment groups had the lower mean levels than the control group for both the black and white probationers. The Asian probationers showed somewhat different patterns — the treatment group had the higher mean level than the control group.

According to Part II in the table, the CBT program was effective in reducing recidivism for the black high-risk probationers when the program effect was measured between the control and treatment groups (ITT). On the other hand, the CBT program effects for the white and Asian high-risk probationers were not statistically meaningful, which may provide the impression that the CBT program would not be effective for those race groups. However, those results should be cautiously interpreted since race was not randomly assigned as a condition of the experiment.

For the white probationers, the Bayesian contrast HDI of λ just marginally included a zero value, while most of the λ estimates fell below zero. Given that a Bayesian HDI can be regarded as a probability distribution, the current result rather indicates that the probability that the CBT program would be effective in reducing the white probationer recidivism is fairly close to 1. Also, the conclusion for the Asians was less reliable because of the too wide 95% HDIs due to small sample sizes, thus warranting further studies on the Asian probationers. Therefore, from the Bayesian perspective, the CBT program was likely effective in reducing recidivism rates for the black or white probationers, or at least it was not conclusive for the Asian probationers.

Meanwhile, when the frequentist NHST model was applied, the CBT program was not effective in reducing probationer recidivism rates at all for all the race subgroups. Note also that all the frequentist 95% confidence intervals were way wider than the corresponding Bayesian 95% HDIs.

Table 3-5. Mean and Intervals of λ Levels and λ Contrast, According to Race Subgroups

	Bayesian Approach		Frequentist Approach	
	Mean	95% HDI of λ (Lower Bound, Upper Bound)	Point Estimation of λ_{MLE}	95% CI of λ_{MLE} (Lower Bound, Upper Bound)
I. Estimates of λ Levels				
I-(1) White				
1. Control Group (N=94)	3.89	(3.45, 4.44)	3.28	(2.22, 4.34)
2. Treatment Group (N=103)	3.36	(2.94, 3.74)	3.12	(1.66, 4.58)
I-(2) Black				
1. Control Group (N=321)	4.62	(4.35, 4.85)	4.31	(2.85, 5.78)
2. Treatment Group (N=322)	4.04	(3.83, 4.30)	3.79	(2.58, 5.00)
I-(3) Asian				
1. Control Group (N=19)	3.28	(2.26, 4.17)	2.74	(0.34, 5.13)
2. Treatment Group (N=21)	4.37	(3.37, 5.36)	4.10	(-0.47, 8.66)
II. Estimates of Contrast, $\lambda_{Treatment} - \lambda_{control}$				
II-(1) White				
2. - 1. Contrast	-0.53	(-1.78, 0.04)	-0.16	(-1.95, 1.63)
II-(2) Black				
2. - 1. Contrast	-0.58	(-0.91, -0.19)*	-0.52	(-2.42, 1.38)
II-(3) Asian				
2. - 1. Contrast	1.09	(-0.17, 2.58)	1.36	(-3.68, 6.40)

Note: * Indicates, from the Bayesian perspective, that the most likely 95% λ contrast estimates between the control and treatment groups do not include zero, and therefore the two groups differ.

indicates, from the frequentist (NHST) perspective, that the null hypothesis that the difference between the control and treatment groups equals zero is rejected at the 95% confidence level.

Table 3.6 reports the CBT effectiveness results for those who had prior probation experiences before the admission to the current probation and those who had not. For those who were new to the probation system, the CBT program was not effective even from the Bayesian perspective. For those who had ever experienced other probation programs before the current one, on the other hand, the CBT program tended to be effective in reducing the recidivism rates of the probationers who had ever experienced other probation programs before. However, when the CBT program was evaluated from the frequentist view, the CBT program was no longer effective at all for both subgroups.

Table 3-6. Mean and Intervals of λ Levels and λ Contrast, According to Prior Probation Experience

	Bayesian Approach		Frequentist Approach	
	Mean	95% HDI of λ (Lower Bound, Upper Bound)	Point Estimation of λ_{MLE}	95% CI of λ_{MLE} (Lower Bound, Upper Bound)
I. Estimates of λ Levels				
I-(1) Not Experienced Group				
1. Control Group (N=167)	4.75	(4.40, 5.12)	4.36	(2.76, 5.95)
2. Treatment Group (N=146)	4.95	(4.60, 5.33)	4.64	(3.05, 6.24)
I-(2) Experienced Group				
1. Control Group (N=280)	4.39	(4.12, 4.65)	4.08	(2.54, 5.61)
2. Treatment Group (N=311)	3.50	(3.29, 3.73)	3.26	(2.10, 4.42)
II. Estimates of Contrast, $\lambda_{Treatment} - \lambda_{control}$				
II-(1) Not Experienced Group				
2. - 1. Contrast	0.20	(-0.29, 0.74)	0.28	(-1.96, 2.53)
II-(2) Experienced Group				
2. - 1. Contrast	-0.88	(-1.21, -0.53)*	-0.81	(-2.73, 1.11)

Note: * Indicates, from the Bayesian perspective, that the most likely 95% λ contrast estimates between the control and treatment groups do not include zero, and therefore the two groups differ.

indicates, from the frequentist (NHST) perspective, that the null hypothesis that the difference between the control and treatment groups equals zero is rejected at the 95% confidence level.

Table 3-7. Mean and Intervals of λ Levels and λ Contrast, According to Ratios of "High" Vote Subgroups

	Bayesian Approach		Frequentist Approach	
	Mean	95% HDI of λ (Lower Bound, Upper Bound)	Point Estimation of λ_{MLE}	95% CI of λ_{MLE} (Lower Bound, Upper Bound)
I. Estimates of λ Levels				
I-(1) Low Ratio of "High" Votes				
1. Control Group (N=235)	3.29	(3.05, 3.54)	3.00	(1.69, 4.31)
2. Treatment Group (N=248)	3.73	(3.50, 4.01)	3.49	(2.06, 4.93)
I-(2) High Ratio of "High" Votes				
1. Control Group (N=212)	5.90	(5.54, 6.25)	5.50	(3.61, 7.38)
2. Treatment Group (N=209)	4.25	(3.95, 4.54)	3.95	(2.80, 5.10)
II. Estimates of Contrast, $\lambda_{Treatment} - \lambda_{control}$				
II-(1) Low Ratio of "High" Votes				
2. - 1. Contrast	0.44	(0.10, 0.80)	0.50	(-1.44, 2.43)
II-(2) High Ratio of "High" Votes				
2. - 1. Contrast	-1.65	(-2.11, -1.19)*	-1.54	(-3.74, 0.65)

Note: * Indicates, from the Bayesian perspective, that the most likely 95% λ contrast estimates between the control and treatment groups do not include zero, and therefore the two groups differ.

indicates, from the frequentist (NHST) perspective, that the null hypothesis that the difference between the control and treatment groups equals zero is rejected at the 95% confidence level.

Table 3.7 reports the CBT program effectiveness for those who had low and high ratios of the Random Forest "high" votes. All the probationers in the current Philadelphia experiment were classified as high-risk persons because their corresponding ratios of Random Forest "high" votes were greater than the ratios of "medium" or "low" votes. However, among those high-risk probationers, some of them were predicted to have higher risks with greater ratios than the others. Therefore, this analysis for the suggested two subgroups actually addresses the question of whether the "risk principle" (Andrews,

Bonta, and Hoge, 1990), that a higher risk should receive a higher level of service, is appropriate.

In Table 3.7, the Bayesian results show that the CBT program was effective in reducing recidivating behaviors of the relatively higher-risk probationers, while it was not effective for the relatively lower-risk probationers. These results support the risk principle. However, the frequentist t-test results suggest that the CBT program was not effective in reducing recidivism rates at all for both subgroups.

3.6. Robustness Test: Different Priors

While Tables 3.3 - 3.7 consistently reported some positive results for the CBT program, the possibility cannot be excluded that the positive results were driven by the choice of the specific Cohen et al.'s prior. As the robustness test to address this concern, Table 3.8 provides the Bayesian estimates of the λ contrasts when the two different priors were used, the Jeffreys (see equation (6) in Section 3.4.3.) and the flat priors (see equation (7) in Section 3.4.3). Even with the different priors, the Bayesian estimate results of the λ contrasts rarely varied. This consistency suggests that the Bayesian CBT program results were not driven by a specific prior choice.

Table 3-8. Robustness Test: Estimates of Contrasts with Different Priors

Estimates of Contrast, $\lambda_{\text{Treatment}} - \lambda_{\text{control}}$	95% HDI of λ (Lower Bound, Upper Bound)		
	Current Reference Prior	Jeffreys Prior	Flat Prior
<u>I. All</u>			
2. - 1. Contrast	(-0.83, -0.25)*	(-0.84, -0.30)*	(-0.83, -0.27)*
<u>II. Age</u>			
II-(1) Age in 10s 2. - 1. Contrast	(-3.81, -1.89)*	(-3.83, -1.96)*	(-3.76, -1.93)*
II-(2) Age in 20s 2. - 1. Contrast	(-0.13, 0.79)	(-0.14, 0.76)	(-0.44, 0.45)
II-(3) Age in 30s 2. - 1. Contrast	(-1.95, -0.89)*	(-1.91, -0.84)*	(-1.88, -0.79)*
II-(4) Age in 40s 2. - 1. Contrast	(-0.77, 0.65)	(-0.76, 0.64)	(-0.72, 0.60)
II-(5) Age in 50s and more 2. - 1. Contrast	(-1.50, 1.37)	(-1.26, 1.59)	(-1.25, 1.36)
<u>III. Race</u>			
III-(1) White 2. - 1. Contrast	(-1.78, 0.04)	(-1.10, 0.09)	(-1.16, 0.04)
III-(2) Black 2. - 1. Contrast	(-0.91, -0.19)*	(-0.90, -0.23)*	(-0.91, -0.22)*
III-(3) Asian 2. - 1. Contrast	(-0.17, 2.58)	(-0.15, 2.48)	(-0.04, 2.50)
<u>IV. Probation Experience</u>			
IV-(1) Not Experienced 2. - 1. Contrast	(-0.29, 0.74)	(-0.38, 0.66)	(-0.33, 0.69)
IV-(2) Experienced 2. - 1. Contrast	(-1.21, -0.53)*	(-1.22, -0.56)*	(-1.22, -0.53)*
<u>V. Level of RF "High" Votes</u>			
VI-(1) Low Level of "High" Votes 2. - 1. Contrast	(0.10, 0.80)	(0.09, 0.83)	(0.08, 0.79)
VI-(2) High Level of "High" Votes 2. - 1. Contrast	(-2.11, -1.19)*	(-2.08, -1.22)*	(-2.09, -1.16)*

Note: * Indicates, from the Bayesian perspective, that the most likely 95% λ contrast estimates between the control and treatment groups do not include zero, and therefore the two groups differ.

3.7. Discussion and Conclusion

According to the current study's results, the cognitive-behavioral therapy (CBT) program is, in general, effective in reducing recidivism among high-risk probationers in terms of the intention-to-treat (ITT) effect. In particular, the current study found that the CBT program is more likely to be effective for some subgroups of probationers; those who were between 10-19 or 30-39 years old at the time they begin their probation case, those who ever experienced probation before they were admitted to the current probation program, and those who had a higher recidivism risk measured by a high ratio of "high" risk prediction votes of the random forest model.

However, the fact that the Philadelphia CBT program was effective for some selected subgroups of probationers does not necessarily mean that the CBT program should be applied to those selected subgroups only, excluding the other probationer subgroups for whom the CBT program was not effective. Rather, the current results need to be understood as a practical guidance for the efficiency purpose in limited situations. The current results may be useful when there is a need to decide to whom the CBT program is applied first, due to budget constraints for program implementation. However, the prioritization is not supposed to violate fundamental values that appropriate correctional interventions have. For example, the prioritization efficiency should not surpass the risk principle that a higher risk should receive a higher level of service (Andrews, Bonta, and Hoge, 1990). Also, it should not conflict with the need to treat all probationers in an appropriately equally fashion. A practical application of the current results for efficiency will be meaningful only when such fundamental values are not

violated. In addition, because subgroups were not randomly assigned as a condition of the experiment, one should be mindful that the evaluation of subgroups is only descriptive.

Importantly, the current paper contributes to the literature on criminal justice evaluation methods. It is notable that, when compared to the frequentist null hypothesis significance testing (NHST)'s t-test, the Bayesian hierarchical model detects more meaningful differences between the CBT (treatment) and non-CBT (control) groups in the current setting. In particular, the Bayesian model frequently detects the effect of the intention-to-treat (ITT), while the frequentist t-test model seldom does. In that sense, the Bayesian approach may result in a different conclusion for two-group comparisons than the frequentist approach (Kruschke, 2011a). This consequence conflicts with the common impression that, for a simple analysis situation such as a two-group comparison, a full Bayesian analysis may be unnecessary because it produces virtually the same results as a frequentist t-test despite its further complexity (for example, Brooks, 2003: 2694).

The differential outcomes between the frequentist and Bayesian analyses seem to be associated with a NHST t-test's insufficient statistical power. In general, as described in the equation (13) in Section 3.4.5., a NHST t-test confidence interval or t-statistic relies heavily on measures of a sample of size N and a sample standard deviation s , along with a point estimation value. When the sample size N decreases or the dispersion of the sample standard deviation s increases, the NHST t-test has less statistical power. For the current evaluation, it seems that the outlier values amplify the sample standard deviations, thus lowering the statistical power of the NHST t-tests and decreasing the likelihood of detecting any meaningful differences between the groups.

Meanwhile, a hierarchical Bayesian model in general has strength in neutralizing outlier effects through the process "shrinkage." As individual Bayesian parameter estimates are calculated as weighted averages between data and hyperparameter values, the estimates shrinking toward hyperparameter values are less influenced by outliers (Kruschke, 2011b: 212). In the current Bayesian hierarchical Gamma Poisson model, the effects of outliers on λ are alleviated because $E(\lambda|\underline{Y})$ is the weighted average of data and the Gamma distribution mean, $\Delta \cdot \underline{Y} + (1 - \Delta) \cdot \mu$ (where $\mu = \alpha\beta$ is the Gamma distribution mean, $\Delta = \frac{\beta}{\beta+1}$ is the weight), implying that λ shrinks toward μ and that data outliers are less influential than when there is no shrinkage. Indeed, while the means of the Bayesian 95% HDIs were similar to the point estimations of the frequentist NHST methods, the widths of the Bayesian HDIs were smaller than those of the NHST t-test confidence intervals.

Of course, these results do not necessarily lead to the argument that the Bayesian approach should always be preferred over the frequentist approach. With a reasonably large sample size and no outliers, the frequentist approach can efficiently yield appropriate conclusions with relatively simple models. Given that the Bayesian approach virtually always requires far more complexity in modeling to reflect parameter uncertainties, along with consideration of a prior selection that often becomes controversial to a skeptical audience, the relative simplicity and efficiency are clearly strengths of frequentist models.

Outliers and small sample sizes are a quite common phenomenon within criminal justice. Given that a relatively small proportion of criminals tend to account for a large

proportion of crimes (Farrington, 1988; Moffitt, 1993; Wolfgang, Figlio, and Sellin, 1972), it is likely that a crime data set will include some units that have outlier outcomes. In addition, data on complex interventions, as in this case, often have limited accessibility by their very nature. For example, data on criminals who are arrested or in the custody of the correctional systems are not fully open to the public. Also, sometimes external factors may influence the scope or quality of the data that can be collected. Though not relevant in this case, experimental data collection may be incomplete due to a lack of funding or the occurrence of unexpected external interruptions. Small sample size is likely to be a problematic issue in such situations.

Given the qualifications, Martinson's (1974) "nothing works" conclusion might be, at least to some extent, attributed to a failure of detecting effects, not a failure of the rehabilitative correctional programs themselves (Lipsey and Cullen, 2007; Weisburd, Lum, Yang, 2003). From the frequentist perspective, a low level of statistical power associated with data quality may have led to a false failure to reject the null hypothesis that a program had no effect, when the program actually had an impact. In the same vein, Rossi's "Iron Law" of evaluation, the principle that "the expected value of any net impact assessment of any large scale social program is zero" (Rossi, 1987), may need to be revisited. As Rossi appropriately indicated, the Iron Law means not a firm natural law but a robust probabilistic tendency. The Iron Law's robustness is based on the past empirical results of intervention programs. The fact that a program evaluation result may differ depending on the evaluation method choice attenuate the robustness of the Iron Law.

Although the current paper's analysis is based on strong Bayesian modeling, it is not free from limitations that should be to be further addressed in future research. With regard to the modeling, partially due to the limited time span of the current data set (limited to one year only), the current hierarchical Gamma Poisson model assumes that the individual recidivism risk parameter, λ_i , is time-invariant. However, this assumption may be unrealistic. Recidivism rates tend to be highest in the first year after release on probation or parole and to decrease thereafter (Petersilia, 2011). Future research needs to consider this temporal recidivism trajectory in models when the experiment results data are gathered over longer periods. One possibility is to use the Bayesian hierarchical regression modeling for the parameter $\lambda_i = \alpha_i + \beta_i \cdot t$, where t is a time variable, and α_i , β_i are individual coefficients.

The current analysis should also be qualified with regard to the operationalization of the dependent variables. In this case, only crimes that lead to arrests recorded within Philadelphia's administrative record system are considered within the outcome variable. This may undercount the extent of criminal behaviors. At the same time, this measure may underestimate the true impact of the program itself. CBT programs can influence non-behavioral factors, including anger or psychological stimulus related to deviance (Lösel, 1995). The current model assumes that the violation outcome can represent non-behavioral effects because non-behavioral factors are implicitly assumed to be one of the causes of those violations. However, for a truly comprehensive evaluation, these outcomes should be assessed separately.

“It also is worth considering the generalizability and limited impact of the current CBT program effects. In an intervention impact pyramid that Frieden (2010) puts forward, the CBT program corresponds to the fifth tier from the bottom of "counseling and education." This tier requires a high level of individual effort. The success of any program in this tier may depend on the program beneficiaries (i.e. probationers) and how much effort these beneficiaries are required to exert in order to receive the benefits of the program. Unlike passive programs that ask little, if anything, of their beneficiaries, active intervention programs like CBT require substantive commitment, engagement, and high levels of maintenance by their beneficiaries, thereby limiting their probability of impact and success (Baker, 1981). Therefore, when the current CBT program is extended to a larger population, the current positive effects may not necessarily hold. One cannot exclude the possibility that the well-sampled high-risk probationers who were ready for change led to the current positive CBT effects. A majority of a larger population may not be ready for change. Even if they are ready when they participate in a CBT program, the program may fail for a large portion of probationers when the CBT program is scaled up to larger populations, outside the study cohort. That is because the success of the program requires a long and costly effort by an individual. A program such as CBT that depends greatly on individual efforts has limited impact generalizability, compared with other programs that focus on more basic structural changes to the contexts and surroundings. Moreover, if left unaddressed, negative or unsafe contextual environments may stymie the success of most individually-focused, lifestyle change programs such as CBT (Branas and MacDonald, 2014; Institute of Medicine, 2003).

Despite these limitations, it is worth reemphasizing this strength of this paper. Practically, by investigating for whom this particular CBT program worked (and for whom it did not), this paper suggests to probation practitioners a way to more efficiently implement similar programs. This is important, as CBT is widely considered to be one of the most promising evidence-based approaches to reducing recidivism.

Also, from a theoretical perspective, this is one of the rare papers that employ a full Bayesian modeling form to evaluate a criminal justice intervention. With a strong Bayesian hierarchical modeling, which is relatively robust for small sample size and outlier problems, this paper provides evidence that the CBT program is more effective than what the common frequentist NHST's approach would suggest. This can be a solid methodological contribution to the criminal justice evaluation field, because many program evaluations in that field are subject to small sample size and outlier problems, which may lead to an artificial inflation of false negative conclusions or failures to reject the null hypothesis of no difference between probation treatment and usual care. In that sense, the Bayesian approach can provide balanced and complementary views for evaluations. Given the strengths and contextual appropriateness, a Bayesian-based approach should be more widely accepted within criminal justice program evaluations.

APPENDIX 3-A. An Exploratory Simulation for a Treatment-on-the-Treated Effect (ETT) Estimation

As mentioned before (in footnote 45, p.108), the formal Bayesian estimation for the treatment-on-the-treated effect (ETT), following Imbens and Rubin (1997) and Long, Little, and Lin (2010) is beyond this paper's purview. Their approaches require an additional variable representing a missing principal compliance status (complier, never-taker, always-taker, and defier) and different modeling using different algorithms. Instead, this appendix section tries to suggest a simple simulation-based ETT estimate range for the current study samples, using the current Monte Carlo grid sampling method.

The current study assumes the complier treatment group to comprise those who were assigned to the CBT-ISP treatment group and attended at least one of the assigned CBT classes. The complier treatment group is likely to be different from the non-complier treatment group in its members' willingness to change behaviors. Comparing the complier treatment group with the control group may lead to a bias because the control group includes not only potential compliers but potential non-compliers. Thus, the low willingness to change behaviors of the control group's potential non-compliers may influence the comparison result. To avoid a bias, the treatment compliers should be compared only to the potential compliers of the control group. However, because the control group never was exposed to the CBT treatment, it is not known whether a control group object belongs to potential compliers or non-compliers.

This section includes two strong assumptions. It is assumed that the control group has the same proportion of compliers as that does the treatment group. Because the current treatment group has 250 compliers out of the 457 treatment group members (54.7%), the control group is assumed to have around 244 potential compliers out of the 447 control group members. In addition, it is assumed that if these potential control group compliers are exposed to the CBT treatment, they would experience a similar proportional recidivism reduction as the treatment group compliers. Table 3-9 reports that the estimated outcome means for the treatment group compliers who were assigned CBT-ISP and attended at least one CBT class are smaller than those for all treatment group members by 5–6% ($= (3.95-3.75)/03.95$ or $(3.70-3.46)/3.70$). Therefore, it is assumed that the potential control group compliers would have 0.94–0.95 times the original frequency outcomes if they were exposed to the CBT treatment.

Table 3-9. Mean and Intervals of λ Levels for All Control, All Treatment, and Treatment Attended Groups

	Bayesian Approach		Frequentist Approach	
	Mean	95% HDI of λ (Lower Bound, Upper Bound)	Point Estimation of λ_{MLE}	95% CI of λ_{MLE} (Lower Bound, Upper Bound)
I. Estimates of λ Levels				
1 All Control Group (N=447)	4.51	(4.30, 4.73)	4.18	(3.05, 5.31)
2. All Treatment Group (N=457)	3.95	(3.78, 4.15)	3.70	(2.77, 4.64)
2.1. Treatment Attended (N=250)	3.75	(3.50, 4.02)	3.46	(2.18, 4.75)

Note: This table is an extension of the upper part of Table 3.3 (p.122) for the Treatment Attended group. The Bayesian estimates are obtained through the Monte Carlo grid sampling method. "Treatment Attended" indicates the treatment group compliers who were assigned CBT-ISP and attended at least one CBT class.

With these assumptions for the potential complier proportion of the control group, the following procedures are conducted:

(3A-A) Among the 447 control groups, 244 samples are randomly selected and are tentatively considered as the potential control group compliers.

Assuming that they are exposed to the CBT treatment, their frequency outcomes are modified to be δ times original amounts. δ , the control group compliers depreciation factor, is set to 0.95 and 0.90.

(3A-B) For a Bayesian 95% HDI for the outcome contrast between the treatment and control group compliers, the Monte Carlo grid sampling (in Section 3.4.2–3.4.6) is applied to the treatment group's 250 compliers and the control group's tentative 244 potential compliers.

(3A-C) For a frequentist 95% confidence interval for the outcome contrast between the treatment and control group compliers, the t-test (in Section 3.4.7) is conducted for the two complier groups.

(3A-D) Repeat the above (3A-A)–(3A-C) procedures 1000 times.

The simulation results are reported in Table 3-10. When the Bayesian Monte Carlo grid method was applied, more than 50% of the simulated 95% HDIs did not include zero. When the control group compliers depreciation factor δ was 0.95, 62.4% of the 95% HDIs were less than zero, which corresponds to the argument that the CBT treatment is effective in reducing recidivism for the high-risk probationers. However, 34.3% of the 95% HDIs included zero. This one third of the simulation results indicates

the possibility that the CBT treatment might have no effect on recidivism reduction. 3.3% of the 95% HDIs were even greater than zero, which means that the CBT treatment might rather promote recidivism.

It is expected that a smaller control group depreciation factor δ is associated with a weaker evidence of the CBT treatment effectiveness. When δ was 0.90, only 48.8% of the 95% HDIs were less than zero. Meanwhile, the proportion of the 95% HDIs that included zero was 44.4%, which was an almost equivalent proportion to that of the 95% HDIs being less than zero. The proportion of the 95% HDIs that were even greater than zero was 6.8%.

On the other hand, when the frequentist t-test method is applied, all the 95% CIs included zero within them. From the frequentist view, the CBT treatment was expected to have no effect on recidivism reduction, even when the effect was estimated in terms of the treatment-on-the-treated effect (ETT).

Table 3-10. Distribution of 1000 simulated 95% Bayesian and Frequentist Intervals

	Bayesian Approach	Frequentist Approach
	Number of 95% HDIs, out of 1,000 simulations	Number of 95% CIs, out of 1,000 simulations
When there are treatment group compliers (N=250) and control group compliers (N=244), contrasts of $\lambda_{\text{Treatment Compliers}} - \lambda_{\text{Control Compliers}}$ are estimated 1,000 times.		
I. When $\delta = 0.95$,		
II-1. 95% interval < 0	624 (=62.4%)	0 (=0%)
II-2. $0 \in$ 95% interval	343 (=34.3%)	1000 (=100%)
II-3. $0 <$ 95% interval	33 (=3.3%)	0 (=0%)
II. When $\delta = 0.90$,		
II-1. 95% interval < 0	488 (=48.8%)	0 (=0%)
II-2. $0 \in$ 95% interval	444 (=44.4%)	1000 (=100%)
II-3. $0 <$ 95% interval	68 (=6.8%)	0 (=0%)

REFERENCES

< References for Chapter 1>

- Alba, Richard, and Victor Nee. 1997. "Rethinking Assimilation Theory for a New Era of Immigration." *International Migration Review* 31:826-874.
- Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies," *Multivariate Behavioral Research* 46: 399-424.
- Bang, Heejung, and James M. Robins. 2005. "Doubly Robust Estimation in Missing Data and Causal Inference Models." *Biometrics* 61(4):962-73.
- Baron, Reuben M., and David A. Kenny. 1986. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):1173-1182.
- Berry, John W. 1997. "Immigration, Acculturation, and Adaptation." *Applied Psychology: An International Review* 46(1):5-68.
- Bersani, Bianca E., Thomas A. Loughran, and Alex R. Piquero. 2013. "Comparing Patterns and Predictors of immigrant Offending among a Sample of Adjudicated Youth," *Journal of Youth Adolescence*. Forthcoming. Retrieved on February 4, 2014, from <http://link.springer.com/article/10.1007%2Fs10964-013-0045-z>.
- Bird, Hector R., Glorisa J. Canino, Mark Davies, Cristiane S. Duarte, Vivian Febo, Rafael Ramirez, Christina Hoven, Judith Wicks, George Musa, and Rolf Loeber. 2006a. "A Study of Disruptive Behavior Disorders in Puerto Rican Youth: I. Background, Design, and Survey Methods." *Journal of the American Academy of Child and Adolescent Psychiatry* 45:1032–1041.
- Bird, Hector R., Mark Davies, and Cristiane S. Duarte, Sa Shen, Rolf Loeber, and Glorisa Canino. 2006b. "A Study of Disruptive Behavior Disorders in Puerto Rican Youth: II. Baseline Prevalence, Comorbidity, and Correlates in Two Sites." *Journal of the American Academy of Child and Adolescent Psychiatry* 45:1042–1053.
- Blau, Peter M. 1977. "A Macrosociological Theory of Social Structure." *American Journal of Sociology* 83(1):26-54.
- Borjas. George J. 2000. "Ethnic Enclaves and Assimilation." *Swedish Economic Policy Review* 7:89-122.

- _____. 2005. "The Labor Market Impact of High-skill Immigration." *National Bureau of Economic Research Working Paper* 11217.
- Brown, Ben, and William Reed Benedict. 2004. "Bullets, Blades, and Being Afraid in Hispanic High Schools: An Exploratory Study of the Presence of Weapons and Fear of Weapon-Associated Victimization Among High School Students in a Border Town." *Crime & Delinquency* 50:372–394.
- Bui, Hoan N., and Ornuma Thongniramol. 2005. "Immigration and Self-reported Delinquency: The Interplay of Immigration, Generations, Gender, Race, and Ethnicity." *Journal of Criminal Justice* 28(2):71–80.
- Buriel, Raymond, Silverio Calzada, and Richard Vasquez. 1982. "The Relationship of Traditional Mexican American Culture to Adjustment and Delinquency among Three Generations of Mexican American Male Adolescents." *Hispanic Journal of Behavioral Sciences* 4(1):41–55.
- Card, David. 2005. "Is the New Immigration Really So Bad?" *National Bureau of Economic Research Working Paper* 11547.
- Card, David, and John E. DiNardo. 2000. "Do Immigrant inflows Lead to Native Outflows?" *National Bureau of Economic Research Working Paper* 7578.
- Carvajal, Scott C., Carrie E. Hanson, Andrea J. Romero, and Karin K. Coyle. 2002. "Behavioural Risk Factors and Protective Factors in Adolescents: A Comparison of Latino and Non-Latino Whites." *Ethnicity & Health* 7(3):181–193.
- Clayton, David, and Michael Hills. 1993. *Statistical Models in Epidemiology*. New York, NY: Oxford University Press.
- Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. 1995. "Statistical Methods for Comparing Regression Coefficients between Models." *American Journal of Sociology* 100: 1261–1293.
- Cook, Philip J. 2009. "Crime Control in the City: A Research-Based Briefing on Public and Private Measures." *Journal of Policy Development and Research* 11(1):53–80.
- Decker, Michael R., Anita Raj, and Jay G. Silverman. 2007. "Sexual Violence Against Adolescent Girls: Influences of Immigration and Acculturation." *Violence Against Women* 13:498–513.
- Dinh, Khanh T., Mark W. Roosa, Jenn-YunTein, and Vera A. Lopez. 2002. "The Relationship between Acculturation and Problem Behavior Proneness in a Hispanic Youth Sample: A Longitudinal Mediation Model." *Journal of Abnormal Child Psychology* 30:295–309.

- Emsley, Richard, Mark Lunt, Andrew Pickles, and Graham Dunn. 2008. "Implementing Double-Robust Estimators of Causal Effects." *The Stata Journal* 8(3):334-353.
- Farley, Reynolds, and Richard Alba. 2002. "The New Second Generation in the United States." *International Migration Review* 36:669-701.
- Finch, Brian K., and William A. Vega. 2003. "Acculturation Stress, Social Support, and Self-Rated Health among Latinos in California." *Journal of Immigrant Health* 5(3):109-117.
- Finch, Brian K., Jason D. Boardman, Bohdan Kolody, and William A. Vega. 2000. "Contextual Effects of Acculturation on Perinatal Substance Exposure among Immigrant and Native-Born Latinas." *Social Science Quarterly* 81(1):421-438.
- Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.
- Gibson, Campbell, and Kay Jung. 2006. "Historical Census Statistics on the Foreign-Born population of the United States: 1850 to 2000." *US Census Bureau Population Division Working Paper No. 81*.
- Gibson, Chris L., and Holly V. Miller. 2010. "Crime and Victimization among Hispanic Adolescents: A Multiple Longitudinal Study of Acculturation and Segmented Assimilation." *A Final Report for the W.E.B. Du Bois Fellowship*. National Institute of Justice
- Gonzales, Nancy A., Julianna Deardorff, Diana Formoso, Alicia Barr, and Manuel Barrera Jr. 2006. "Family Mediators of the Relation Between Acculturation and Adolescent Mental Health." *Family Relations* 55(3):318-330.
- Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York, NY: Oxford University Press.
- Gottfredson Michael R., and Travis Hirschi. 1990. *A General Theory of Crime*. Palo Alto, CA: Stanford University Press.
- Greenland, Sander. 1987. "Interpretation and Choice of Effect Measures in Epidemiologic Analyses." *American Journal of Epidemiology* 125(5): 761-768.
- Guo, Shenyang Y., and Mark W. Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage Publications
- Hindelang, Michael J. 1976. *Criminal Victimization in Eight American Cities: A Descriptive Analysis of Common Theft and Assault*. Cambridge, MA: Ballinger Publishing Co.

- Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley, CA: University of California Press.
- Horowitz, Carl F. 2001. *An Examination of US Immigration Policy and Serious Crime*. Washington, DC: Center for Immigration Studies. Retrieved on January 11, 2014, from <http://www.cis.org/articles/2001/crime/toc.html>
- Joffe, Marshall M., Thomas R. Ten Have, Harold I. Feldman, and Stephen E. Kimmel. 2004. "Model Selection, Confounder Control, and Marginal Structural Models: Review and New Applications." *The American Statistician* 58(4): 272-279.
- Johnson, Kenneth M., and Daniel T. Lichter. 2008. "Population Growth in New Hispanic Destinations." *Policy Brief* 8., Carsey Institute
- Lauritsen, Janet L., John H. Laub, and Robert J. Sampson. 1992. "Conventional and Delinquent Activities: Implications for the Prevention of Violent Victimization among Adolescents." *Violence and Victims* 7(2):91-108.
- Lauritsen, Janet L., Robert J. Sampson, and John H. Laub. 1991. "The Link between Offending and Victimization among Adolescents." *Criminology* 29:265-292.
- Lee Matthew T., and Ramino Martinez Jr. 2009. "Immigration Reduces Crime: An Emerging Scholarly Consensus," In William F. McDonald (Ed.), *Immigration, Crime and Justice (Sociology of Crime Law and Deviance)* (Volume 13, pp. 3-16). Bingley, UK: Emerald Group Publishing Limited.
- Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data, 2nd Edition*. New York, NY: Wiley-Interscience.
- Lunceford, Jared K., and Marie Davidian. 2004. "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study." *Statistics in Medicine* 23:2937-2960.
- MacDonald John M., and Jessica Saunders. 2012. "Are Immigrant Youth Less Violent? Specifying the Reasons and Mechanism." *Annals of the American Academy of Political and Social Science* 641:125-147.
- MacDonald, John M., Ricky N. Bluthenthal, Daniela Golinelli, Aaron Kofner, Robert J. Stokes, Amber Sehgal, Terry Fain, and Leo Beletsky. 2009. *Neighborhood Effects on Crime and Youth Violence: The Role of Business Improvement Districts in Los Angeles*. Santa Monica, CA: RAND Co.
- Morenoff, Jeffrey D., and Avraham Astor. 2006. "Immigrant Assimilation and Crime: Generational Differences in Youth Violence in Chicago." In Ramiro Martinez Jr., and Abel Valenzuela Jr. (Eds.), *Immigration and Crime: Race, Ethnicity, and Violence*, (pp. 36-63). New York and London: New York University Press.

- Morenoff, Jeffrey D., Robert J. Sampson, and Stephen W. Raudenbush. 2001. "Neighborhood inequality, collective efficacy and the spatial dynamics of urban homicide." *Criminology* 39:517-559.
- Ousey, Graham C., and Charis E. Kubrin. 2009. "Exploring the Connection between Immigration and Violent Crime Rates in US Cities, 1980-2000." *Social Problems* 56(3): 447-448.
- Palloni, Alberto, and Elizabeth Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." *Demography* 41(3):385-415.
- Palloni, Alberto, and Jeffrey D. Morenoff. 2001. "Interpreting the Paradoxical in the Hispanic Paradox: Demographic and Epidemiologic Approaches." *Annals of the American Academy of Political and Social Science* 954:140-74.
- Park, Robert E. 1936a. "Human ecology." *American Journal of Sociology* 42:1-15.
- _____. 1936b. "Succession, an Ecological Concept." *American Sociological Review* 1:171-179.
- Paternoster, Raymond, Robert Brame, Paul Mazerolle, and Alex Piquero. 1998. "Using the Correct Statistical Test for the Equality of Regression Coefficients." *Criminology* 36(4): 859-866.
- Perlmann, Joel, and Roger Waldinger. 1997. "Second Generation Decline? Children of Immigrants, Past and Present: A Reconsideration." *International Migration Review* 31:893-922.
- Piquero, Alex, and Matthew J. Hickman. 2003. "Extending Tittle's Control Balance Theory to Account for Victimization." *Criminal Justice and Behavior* 30(3):282-301.
- Portes, Alejandro, and Ruben G. Rumbaut. 1996. *Immigrant America: A Portrait*, 2nd Edition. Berkeley, CA: University of California Press.
- _____. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley, CA: University of California Press.
- Portes, Alejandro, and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and Its Variants." *Annals of the American Academy of Political and Social Science* 530:74-96
- Pratt, Travis C., and Francis T. Cullen. 2000. "The Empirical Status of Gottfredson and Hirschi's General Theory of Crime: A Meta-analysis." *Criminology*, 38: 931-964.
- Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods: 2nd Edition*. Thousand Oaks, CA: Sage Publications.
- Redfield, Robert, Ralph Linton, and Melville J. Herskovits. 1936. "Memorandum for the Study of Acculturation." *American Anthropologist* 38:149-152.

- Reid, Lesley W., Harald E. Weiss, Robert M. Adelman, and Charles Jaret. 2005. "The Immigration-Crime Relationship: Evidence across US Metropolitan Areas." *Social Science Research* 34:757-780.
- Rosenbaum Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology* 66: 688-701.
- _____. 2001. "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services and Outcomes Research Methodology* 2: 169-188.
- Rumbaut, Ruben G. 2004. "Ages, Life Stages, and Generational Cohorts: Decomposing the Immigrant First and Second Generations in the United States." *International Migration Review* 38(3):1160-1205.
- _____. 2008(Aug. 21-22). "Undocumented Immigration and Rates of Crime and Imprisonment: Popular Myths and Empirical Realities." Paper presented to the Police Foundation National Conference on "The Role of Local Police: Striking a Balance between Immigration Enforcement and Civil Liberties" in Washington, DC. Retrieved on June 1, 2012, from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1877365.
- Samaniego, Roxana Y., and Nancy A. Gonzales. 1999. "Multiple Mediators of the Effects of Acculturation Status on Delinquency for Mexican American Adolescents." *American Journal of Community Psychology* 27:189-210.
- Sampson, Robert J. 2006 (Mar. 11). "Open Doors Don't Invite Criminals." *New York Times* article. Retrieved on June 1, 2012, from <http://www.nytimes.com/2006/03/11/opinion/11sampson.html>.
- _____. 2008. "Rethinking Crime and Immigration." *Contexts* 7(1): 28-33.
- Sampson, Robert J., and W. Byron Groves. 1989. "Community Structure and Crime: Testing Social Disorganization Theory." *American Journal of Sociology* 94:774-802.
- Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing 'neighborhood effects': Social processes and new directions in research." *Annual Review of Sociology* 28:443-478.
- Sampson, Robert J., Stephen Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918-24.
- Sanderson, Maureen, Ann L. Coker, Robert E. Roberts, Susan R. Tortolero, and Belinda M. Reininger. 2004. "Acculturation, Ethnic Identity, and Dating

- Violence among Latino Ninth-Grade Students." *Preventive Medicine* 39:373–383.
- Scribner, Richard. 1996. "Paradox as Paradigm - The Health Outcomes of Mexican Americans." *American Journal of Public Health* 86(3):303-305.
- Schreck, Christopher J. 1999. "Criminal Victimization and Low Self-Control: An Extension and Test of a General Theory of Crime." *Justice Quarterly* 16:633-654.
- Schreck, Christopher J., Richard A. Wright, and J. Mitchell Miller. 2002. "A Study of Individual and Situational Antecedents of Violent Victimization." *Justice Quarterly* 19:159-180.
- Shaw, Clifford R., and Henry D. McKay. 1942. *Juvenile Delinquency and Urban Areas*. Chicago, IL: University of Chicago.
- Silverman, Jay G., Michele R. Decker, Anita Raj. 2007. "Immigration-based Disparities in Adolescent Girls' Vulnerability to Dating Violence." *Maternal and Child Health Journal* 11(1):37-43.
- Singh, Gopal K., and Mohammad Siahpush, 2002. "Ethnic-Immigrant Differentials in Health Behaviors, Morbidity, and Cause-Specific Mortality in the United States: An Analysis of Two National Data Bases." *Human Biology* 74(1):83-109.
- Smith, Emilie P., and Nancy G. Guerra. 2006. "Introduction." In Nancy G. Guerra, and Emilie .P. Smith (Eds.), *Preventing Youth Violence in a Multicultural Society*, (pp. 3-14). Washington DC: American Psychological Association.
- Smokowski, Paul R., and Martica L. Bacallao. 2006. "Acculturation and Aggression in Latino Adolescents: A Structural Model Focusing on Cultural Risk Factors and Assets." *Journal of Abnormal Child Psychology* 34(5):657-671.
- _____. 2007. "Acculturation, Internalizing Mental Health Symptoms, and Self-esteem: Cultural Experiences of Latino Adolescents in North Carolina." *Child Psychiatry and Human Development* 37(3):273-292.
- Smokowski, Paul R., Corinne Dvaid-Ferdon, and Nancy Stroupe. 2009. "Acculturation and Violence in Minority Adolescents: A Review of the Empirical Literature." *Journal of Primary Prevention* 30:215-263.
- Sommers, Ira, Jeffrey Fagan, and Deborah Baskin. 1993. "Sociocultural influences on the explanation of delinquency for Puerto Rican youths." *Hispanic Journal of Behavioral Sciences* 15:36–62.
- Stock, James H., and Mark W. Watson. 2007. *Introduction to Econometrics: 2nd edition*. Boston, MA: The Pearson Education.
- Stowell, Jacob I., Steven F. Messner, Kelly F. McGeever, and Lawrence E. Raffalovich. 2009. "Immigration and the Recent Violent Crime Drop in the United States:

- A Pooled, Cross-Sectional Time-series Analysis of Metropolitan Areas." *Criminology* 47(3): 889-928.
- U.S. Census Bureau. 2007. *Minority Population Tops 100 Million*. Retrieved on December 5, 2012, from US Census homepage, <http://www.census.gov/newsroom/releases/archives/population/cb07-70.html>.
- Vega, William A., and Andres G. Gil, 1998. "Acculturative Stress and Drug Use Behavior among Immigrant and U.S. Born Latino Adolescents: Toward an Integrated Model." *Drug and Society* 14:55-71.
- Vega, William A., Andres G. Gil, George J. Warheit, Rick S. Zimmerman, and Eleni Apospori. 1993. "Acculturation and Delinquent Behavior among Cuban American Adolescents: Toward an Empirical Model." *American Journal of Community Psychology* 21(1): 113-125.
- Waters, Mary, and Tomas R. Jimenez. 2005. "Assessing Immigrant Assimilation: New Empirical and Theoretical Challenges." *Annual review of Sociology* 31: 105-25.
- Ye, Yu, and Lee Ann Kaskutas. 2009. "Using Propensity Scores to Adjust for Selection Bias When Assessing the Effectiveness of Alcoholics Anonymous in Observational Studies." *Drug and Alcohol Dependence* 104: 56-64.
- Yu, Stella M., Zhihuan J. Huang, Renee H. Schwalberg, Mary Overpeck, and Michael D. Kogan. 2003. "Acculturation and the Health and Well-being of U.S. Immigrant Adolescents." *Journal of Adolescent Health* 33:479-488.
- Zhou, Min. 1997a. "Segmented Assimilation: Issues, Controversies, and Recent Research on the New Second Generation." *International Migration Review* 31: 975-1008.
- _____. 1997b. "Growing Up American: The Challenge Confronting Immigrant Children and Children of Immigrants." *Annual Review of Sociology* 23: 63-95.
- _____. 2001. "Contemporary Immigration and the Dynamics of Race and Ethnicity." In Neil J. Smelser, William J. Wilson, and Faith Mitchell (Eds.), *America Becoming: Racial Trends and Their Consequences*, (pp. 200-242). Washington, DC: National Academy Press.
- Zhou, Min, and Carl L. Bankston III. 1998. *Growing Up American: How Vietnamese Children Adapt to Life in the United States*. New York, NY: Russell Sage Foundation.
- _____. 2006. "Delinquency and Acculturation in the Twenty-first Century: A Decade's Change in a Vietnamese American Community." In Ramiro Martinez Jr., and Abel Valenzuela Jr. (Eds.), *Immigration and Crime: Race,*

Ethnicity, and Violence, (pp. 117-139). New York and London: New York University Press.

< References for Chapter 2>

- Barnes, Geoffrey C. 1995. "Defining and Optimizing Displacement." *Crime and Prevention Studies* 4: 95-113.
- Barr, Robert, and Ken Pease. 1990. "Crime Displacement, Displacement, and Deflection." *Crime and Justice: A Review of Research* 12: 277-318.
- Branas, Charles C., Rose A. Cheney, John M. MacDonald, Vicky W. Tam, Tara D. Jackson, and Thomas R. Ten Have. 2011. "A Difference-in-Differences Analysis of Health, Safety, and Greening Vacant Urban Space." *American Journal of Epidemiology* 174(11): 1296-1306.
- Branas, Charles C., Michael R. Elliott, Therese S. Richmond, Dennis P. Culhane, and Douglas J. Wiebe. 2009. "Alcohol Consumption, Alcohol Outlets, and the Risk of Being Assaulted with a Gun." *Alcoholism: Clinical and Experimental Research* 11(4): 906-915.
- Brantingham, Patricia L., and Paul J. Brantingham. 1993. "Environment, Routine, and Situation: Toward a Pattern Theory of Crime." In *Routine Activity and Rational Choice*, edited by Ronald V. Clarke and Marcus Felson, 259-294. New Brunswick, NJ: Transaction.
- _____. 1995. "Criminality of Place: Crime Generators and Crime Attractors." *European Journal on Criminal Policy and Research* 3(3): 5-26.
- Card, David, and Gordon Dahl. 2011. "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior." *Quarterly Journal of Economics* 126 (1): 103-143.
- Carpenter, Christopher. 2005. "Heavy Alcohol Use and the Commission of Nuisance Crime: Evidence from Underage Drunk Driving Laws." *American Economic Review Papers and Proceedings* 95(2): 267-272.
- Carpenter, Christopher, and Carlos Dobkin. 2009. "The Effect of Alcohol Access on Consumption and Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age." *American Economic Journal Applied Economics*, 1(1): 164-182.
- _____. 2010. "The Drinking Age, Alcohol Consumption, and Crime." Working Paper. http://web.gsm.uci.edu/~kittc/Carpenter_Dobkin_Crime_website_01192011.pdf, accessed June 5, 2013.

- _____. 2011. "Alcohol Regulation and Crime." In *Controlling Crime: Strategies and Tradeoffs*, edited by Philip J. Cook, Jens Ludwig, and Justin McCrary, 291-329. National Bureau of Economic Research, Inc. Chicago, IL: University of Chicago Press.
- Carpenter, Christopher, and Daniel Eisenberg. 2009. "The Effect of Sunday Sales Restrictions on Overall and Day-Specific Alcohol Consumption: Evidence from Canada." *Journal of Studies on Alcohol and Drugs* 70(1): 126-133.
- Carpenter, Christopher, Deborah D. Kloska, Patrick O'Malley, and Lloyd Johnston. 2007. "Alcohol Control Policies and Youth Alcohol Consumption: Evidence from 28 Years of Monitoring the Future." *The B.E. Journal of Economic Analysis & Policy* 7(1): 1-21.
- Clarke, Ronald V., and Patricia Mayhew. 1998. "Preventing Crime in Parking Lots: What We Know and Need To Know." In *Reducing Crime Through Real Estate Development and Planning*, edited by Marcus Felson and Richard B. Peiser. Washington, D.C.: Urban Land Institute.
- Cohen, Mark A., and Alex. R. Piquero. 2009. "New Evidence on the Monetary Value of Saving a High Risk Youth." *Journal of Quantitative Criminology* 25(1): 25-49.
- Cook, Philip J. 2007. *Paying the Tab: The Cost and Benefits of Alcohol Control*. Princeton, NJ: Princeton University Press.
- Cook, Philip J, and Christine P. Durrance. 2013. "The Virtuous Tax: Lifesaving and Crime-Prevention Effects of the 1991 Federal Alcohol-Tax Increase." *Journal of Health Economics* 32: 261-267.
- Cornish, Derek, and Ronald V. Clarke. 1986. "Situational Prevention: Displacement of Crime and Rational Choice Theory." In *Situational Crime Prevention: From Theory into Practice*, edited by Kevin Heal and Gloria Laycock. London, UK: Her Majesty's Stationery Office.
- _____. 1987. "Understanding Crime Displacement: An Application of Rational Choice Theory." *Criminology* 25(4): 933-948.
- DeNavas-Walt, Carmen, Bernadette D. Proctor, and Jessica C. Smith. 2013. *Income, Poverty, and Health Insurance Coverage in the United States: 2012*. U.S. Census Bureau. Washington, DC: U.S. Government Printing Office.
- Department of Transportation of City of Sacramento. 2006. *City of Sacramento Pedestrian Master Plan: Making Sacramento the Walking Capital*. Final Report. http://www.cityofsacramento.org/transportation/dot_media/street_media/sac-ped-plan_9-06.pdf, accessed June 5, 2013.
- Distilled Spirits Council of the United States. "Sunday Alcohol Sales." <http://www.discus.org/policy/sunday/>, accessed June 5, 2013.

- Eck, John. 1993. "The Threat of Crime Displacement." *Problem Solving Quarterly* 6(3): 1-2.
- Gorman, Dennis M., Paul W. Speer, Paul Gruenewald, and Erich W. Labouvie. 2001. "Spatial Dynamics of Alcohol Availability, Neighborhood Structure and Violent Crime." *Journal of Studies on Alcohol* 62(5): 628-636.
- Gorman, Dennis M., Paul W. Speer, Erich W. Labouvie, and Apana P. Subaiya. 1998. "Risk of Assaultive Violence and Alcohol Availability in New Jersey." *American Journal of Public Health* 88(1): 97-100.
- Greenfield, Lawrence A. 1998. *Alcohol and Crime: An Analysis of National Data on the Prevalence of Alcohol Involvement in Crime*. Prepared for the Assistant Attorney General's National Symposium on Alcohol Abuse and Crime. Washington DC: U.S. Department of Justice, Bureau of Justice Statistics.
- Grossman, Michael, Douglas Coate, and Gregory Arluck. 1987. "Price Sensitivity of Alcohol Beverages in the United States." In *Control Issues in Alcohol Abuse Prevention: Strategies for States and Communities*, edited by Harold Holder, 169-198. Greenwich: JAI Press.
- Gruenewald, Paul. 2007. "The Spatial Ecology of Alcohol Problems: Niche Theory and Assortative Drinking," *Addiction* 102: 870-878.
- Gruenewald, Paul, Bridget Freisthler, Lillian Remer, Elizabeth LaScala, and Andrew Treno. 2006. "Ecological Models of Alcohol Outlets and Violent Assaults: Crime Potentials and Geospatial Analysis." *Addiction* 101: 666-677.
- Gruenewald, Paul, Alexander B. Millar, and Peter Roeper, 1996. "Access to Alcohol: Geography and Prevention for Local Communities." *Alcohol Health & Research World* 20(4): 244-251.
- Hakim Simon, and George F. Rengert. 1981. "Introduction." In *Crime Spillover*, edited by Simon Hakim and George F. Rengert. Beverly Hills, CA: Sage.
- Heaton, Paul. 2012. "Sunday Liquor Laws and Crime." *Journal of Public Economics* 96(1-2): 42-52.
- Jacob, Brian A., Lars Lefgren, and Enrico Moretti, 2007. "The Dynamics of Criminal Behavior: Evidence from Weather Shocks." *Journal of Human Resources* 42(3): 489-527.
- Kurland, Justin, Shane D. Johnson, and Nick Tilley. 2013. "Offenses Around Stadiums: A Natural Experiment on Crime Attraction and Generation." *Journal of Research in Crime and Delinquency* 00(0): 1-24. (published online before print: <http://jrc.sagepub.com/content/early/2013/01/11/0022427812471349.full.pdf+html>)

- Lovenheim, Michael F., and Daniel P. Steefel. 2011. "Do Blue Laws Save Lives? The Effects of Sunday Sales Bans on Fatal Vehicle Accidents." *Journal of Policy Analysis and Management* 30(4): 798-820.
- Maloney, Michael T., and Jackson C. Rudbeck. 2009. "The Outcome from Legalizing Sunday Packaged Alcohol Sales on Traffic Accidents in New Mexico." *Accident Analysis and Prevention* 41(5): 1094-1098.
- McMillan, Garnett, and Sandra Lapham. 2006. "Legalized Sunday Packaged Alcohol Sales and Alcohol Related Traffic Crashes and Crash Fatalities in New Mexico." *American Journal of Public Health* 96: 1944-1948.
- Miller, Ted R., David T. Levy, Mark A. Cohen, and Kenya L. C. Cox. 2006. "Costs of Alcohol and Drug-Involved Crime." *Prevention Science* 7(4): 333-342.
- Norström, Thor, and Ole-Jorgen Skog. 2003. "Saturday Opening of Alcohol Retail Shops in Sweden: An Impact Analysis." *Journal of Studies on Alcohol* 64: 393-401.
- Olsson, Orvar, and Per-Olof. H. Wikström. 1982. "Effects of the Experimental Saturday Closing of Liquor Retail Stores in Sweden." *Contemporary Drug Problems* 11: 325-353.
- Rand, Michael R., William J. Sabol, Michael Sinclair, and Howard N. Snyder. 2010. "Alcohol and Crime: Data from 2002 to 2008." Bureau of Justice Statistics Homepage (<http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2313>).
- Rees, Daniel, and Kevin Schnepel. 2009. "College Football Games and Crime." *Journal of Sports Economics* 10:68-87.
- Reppetto, Thomas A. 1976. "Crime Prevention and the Displacement Phenomenon." *Crime & Delinquency* 22:166-177.
- Scribner, Richard, Deborah Cohen, Stephen Kaplan, and Susan H. Allen. 1999. "Alcohol Availability and Homicide in New Orleans: Conceptual Considerations for Small Area Analysis of the Effect of Alcohol Outlet Density." *Journal of Studies on Alcohol* 60(3): 310-316.
- Scribner, Richard, David Mackinnon, and James Dwyer. 1995. "The Risk of Assaultive Violence and Alcohol Availability in Los Angeles County." *American Journal of Public Health* 85(3): 335-340.
- Stehr, Mark. 2007. "The Effect of Sunday Sales Bans and Excise Taxes on Drinking and Cross-Border Shopping for Alcoholic Beverages." *National Tax Journal* 60: 85-105.
- _____. 2010. "The Effect of Sunday Sales of Alcohol on Highway Crash Fatalities." *B.E. Journal of Economic Analysis & Policy* 10(1): 1-20.

Teh, Bing-Ru. 2008. "Do Liquor Stores Increase Crime and Urban Decay? Evidence from Los Angeles." Unpublished Ph.D. Dissertation Paper, University of California, Berkeley

< References for Chapter 3>

- Andrews, D. A., James Bonta, and R. D. Hoge. 1990. "Classification for Effective Rehabilitation: Rediscovering Psychology," *Criminal Justice and Behavior* 17(1): 19-52.
- Anwar, Shamena, and Thomas A. Loughran. 2011. "Testing a Bayesian Learning Theory of Deterrence among Serious Juvenile Offenders," *Criminology* 49: 667-698.
- Baker, Susan P. 1981. "Childhood Injuries: The Community Approach to Prevention," *Journal of Public Health Policy* 2(3): 235-246.
- Barnes, Jeffrey C., and Jordan M. Hyatt. 2012. "Classifying Adult Probationers by Forecasting Future Offending," *Final Technical Report*, National Institute of Justice, U.S. Department of Justice.
(<https://www.ncjrs.gov/pdffiles1/nij/grants/238082.pdf>)
- Bartell, Ted, and L. Thomas Winfree Jr. 1977. "Recidivist Impacts of Differential Sentencing Practices for Felony Offenders," *Criminology* 15(3): 387-396.
- Berger, James. 2006. "The Case for Objective Bayesian Analysis," *Bayesian Analysis* 1: 385-402.
- Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western. 1992a. "A Bayesian Analysis of the Colorado Springs Spouse Abuse Experiment," *The Journal of Criminal Law and Criminology* 83(1): 170-200.
- _____. 1992b. "The Deterrent Effect of Arrest in Incidents of Domestic Violence: A Bayesian Analysis of Four Field Experiments," *American Sociological Review* 57(5): 698-708.
- Berk, Richard A., Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz, Lindsay Ahlman. 2009. "Forecasting Murder within a Population of Probationers and Parolees: a High Stakes Application of Statistical Learning," *Journal of the Royal Statistical Society* 172: 971-1008.
- Bernardo, Jose M. 1979. "Reference Posterior Distribution for Bayesian Inference," *Journal of the Royal Statistical Society Series B Methodological* 41(2): 113-147.

- Branas, Charles C., and John M. MacDonald. 2014. "A Simple Strategy to Transform Health, All Over the Place," *Journal of Public Health Management & Practice* 20(2): 157-159.
- Brooks, Stephen P. 2003. "Bayesian Computation: A Statistical Revolution," *Philosophical Transactions of the Royal Society of London: Series A. Mathematical, Physical and Engineering Sciences* 361: 2681-2697.
- Caldwell, Morris G. 1951. "Preview of a New Type of Probation Study Made in Alabama," *Federal Probation* 15(2): 3-11.
- Cockerill, R. W. 1975. "Probation Effectiveness in Alberta," *Canadian Journal of Criminology and Corrections* 17(4): 284-291.
- Cohen, Jacob. 1994. "The Earth is Round ($p < .05$)," *American Psychologist* 49:997-1003.
- Cohen, Jacqueline, Daniel Nagin, Garrick Wallstrom, and Larry Wasserman. 1998. "Hierarchical Bayesian Analysis of Arrest Rates," *Journal of the American Statistical Association* 93(44): 1260-1270.
- Cunniff, Mark A. 1987. "A Sentencing Postscript: Felony Probationers under Supervision in the Community, 1983." *ISPSR* 8647. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Davis, George F. 1964. "A Study of Adult Probation Violation Rates by Means of the Cohort Approach," *Journal of Criminal Law, Criminology and Police Science* 55(1): 70-85.
- Deller, Steven C., Lindsay Amiel, Melissa Deller. 2011. "Model Uncertainty in Ecological Criminology: An Application of Bayesian Model Averaging With Rural Crime Data," *International Journal of Criminology and Sociological Theory* 4(2): 683-717.
- Efron, Bradley, and D. Feldman. 1991. "Compliance as an Explanatory Variable in Clinical Trials," *Journal of the American Statistical Association* 86(413): 9-26.
- Howard, George S., Scott E. Maxwell, Kevin J. Fleming. 2000. "The Proof of the Pudding: an Illustration of the Relative Strengths of Null Hypothesis, Meta-Analysis, and Bayesian Analysis," *Psychological Methods* 5(3):315-332.
- Institute of Medicine. 2003. *The Future of the Public's Health in the 21st Century*. Washington, DC: National Academies Press.
- Hyatt, Jordan M. 2013. *The Impact of Cognitive Behavioral Therapy on the Recidivism of High Risk Probationers: Results from a Randomized Trial*. (doctoral dissertation). University of Pennsylvania, Philadelphia, PA.
- Farrington, David P. 1986. "Age and Crime," *Crime and Justice* 7: 189-250.
- _____. 1988. "Predictors, Causes, and Correlates of Male Youth Violence," *Crime and Justice* 24: 421-475.

- Forcina, Antonio. 2006. "Causal Effects in the Presence of Noncompliance: A Latent Variable Interpretation," *Merton - International Journal of Statistics* 64(3): 275-301.
- Frieden, Thomas R. 2010. "A Framework for Public Health Action: The Health Impact Pyramid," *American Journal of Public Health* 100(4): 590-595.
- Frölich, Marcus, and Blaise Melly. 2013. "Identification of Treatment Effects on the Treated with One-sided non-Compliance," *Econometric Reviews* 32(3): 384-414.
- General Accounting Office. 1976. "State and County Probation: Systems in Crisis," *Comptroller General's Report to the Congress*. Washington, DC: General Accounting Office.
- Gelman, Andrew. 2003. "A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-Fit Testing," *International Statistical Review* 71(2): 369-382.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Geman, Donald, and Stuart Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6): 721-741.
- General Accounting Office. 1976. "State and County Probation Systems in Crisis." Washington, DC: U.S. Government Printing Office.
- Glaze, Lauren E. 2011. "Correctional Populations in the United States, 2010." *BJS Bulletin*. Washington, DC: Bureau of Justice Statistics.
- Gross, Deborah, and Louis Fogg. 2004. "A Critical Analysis of the Intent-to-Treat Principle in Prevention Research," *The Journal of Primary Prevention* 25(4): 475-489.
- Imbens, Guido W., and Donald B. Rubin. 1997. "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance," *The Annals of Statistics* 25(1): 305-327.
- Jeffreys, Harold. 1967. *Theory of Probability*. London, UK: Oxford University Press.
- Kruschke, John K. 2011a. "Bayesian Estimation Supersedes the t Test," *Journal of Experimental Psychology: General* 142(2): 573-603.
- _____. 2011b. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Elsevier.
- Landenberger, Nana A., and Mark W. Lipsey. 2005. "The Positive Effects of Cognitive-Behavioral Programs for Offenders: A Meta-analysis of Factors Associated with Effective Treatment," *Journal of Experimental Criminology* 1: 451-476.

- Landis, Judson R., James D. Mercer, and Carole E. Wolff. 1969. "Success and Failure of Adult Probationers in California," *Journal of Research in Crime and Delinquency* 6(1): 34-40.
- Langan, Patrick A., and Mark A. Cuniff. 1992. "Recidivism of Felons on Probation, 1986-89," *Special Report*. Washington D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Langan, Patrick A., and David J. Levin. 2002. "Recidivism of Prisoners Released in 1994," *Special Report*. Washington D.C.: U.S. Department of Justice, Bureau of Justice Statistics.
- Law, Jane, Matthew Quick, and Ping Chan. 2014. "Bayesian Spatio-Temporal Modeling for Analysing Local Patterns of Crime over Time at the Small-Area level," *Journal of Quantitative Criminology* 30: 57-78.
- Liberton, Michael, Mitchell Silverman, and William R. Blount. 1992. "Predicting Probation Success for the First-Time Offender," *International Journal of Offender Therapy and Comparative Criminology* 36(4): 355-347.
- Lipsey, Mark W. 1992. "Juvenile Delinquency Treatment: A Meta-analytic Inquiry into Variability of Effects," In Thomas D. Cook, Harris Cooper, et al. (Eds.) *Meta-analysis for Explanation*: 83-127. New York, NY: Russell Sage Foundation.
- _____. 1995. "What do We Learn from 400 Research Studies on the Effectiveness of Treatment with Juvenile Delinquents," In James McGuire, *What Works: Reducing Reoffending, Guidelines from Research and Practice*: 62-78. Chichester, UK: John Wiley and Sons.
- _____. 2009. "The Primary Factors that Characterize Effective Interventions with Juvenile Offenders: A Meta-analytic Overview," *Victims and Offenders: An International Journal of Evidence-based Research, Policy, and Practice* 4: 124-147.
- Lipsey, Mark W., and Francis T. Cullen. 2007. "The Effectiveness of Correctional Rehabilitation: A Review of Systematic Reviews," *Annual Review of Law and Social Science* 3: 297-320.
- Lipsey Mark W., and Nana A. Landenberger. 2005. "Cognitive-Behavioral Interventions: A Meta-Analysis of Randomized Controlled Studies," in B.C. Walsh and D. P. Farrington, *Preventing Crime: What works for Children, Offenders, Victims, and Places* (pp. 57-71). Berlin, Germany: Springer.
- Lipsey Mark W., Gabrielle L. Chapman, and Nana A. Landenberger. 2001. "Cognitive-Behavioral Programs for Offenders," *Annals of the American Academy of Political and Social Science* 578: 144-157.

- Lipsey, Mark W., Nana A. Landenberger, and Sandra J. Wilson. 2007. "Effects of Cognitive-Behavioral Programs for Criminal Offenders." *Campbell Systematic Reviews* 207:6. Campbell Collaboration.
- Long, Qi, Roderick J. A. Little, and Xihong Lin. 2010. "Estimating Causal Effects in Trials Involving Multitreatment Arms Subject to Non-Compliance: A Bayesian Framework," *Applied Statistics* 59(3): 513-531.
- Lösel, Friedrich. 1995. "The Efficacy of Correctional Treatment: A Review and Synthesis of meta-evaluations," In James McGuire, *What Works: Reducing Reoffending, Guidelines from Research and Practice*: 79-111. Chichester, UK: John Wiley and Sons.
- Lowenkamp, Christopher T., Edward J. Latessa, and Alexander M. Holsinger. 2006. "The Risk Principal in Action: What Have We Learned from 13,676 Offenders and 97 Correctional Programs," *Crime and Delinquency* 52(1): 77-93.
- Martin, Ronald L., C. Robert Cloninger, and Samuel B. Guze. 1978. "Female Criminology and Prediction of Recidivism," *Archives of General Psychiatry* 35(2): 207-214.
- Martinson, Robert. 1974. "What Works? Questions and Answers about Prison Reform," *The Public Interest* 35: 22-54.
- McCarthy, Belinda and Robert Langworthy. 1987. *Older Offenders: Perspectives in Criminology and Criminal Justice*. New York, NY: Praeger.
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics* 21(6): 1087-1092.
- Miller, Robert B. 1980. "Bayesian Analysis of the Two-Parameter Gamma Distribution," *Technometrics*, 22(1): 65-69.
- Moffitt, Terrie E. 1993. "Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy," *Psychological Review* 100: 674-701.
- Monahan, John. 1981. *Predicting violent behavior: An Assessment of Clinical Techniques*. Beverly Hills, CA: Sage Publications.
- Morgan, Kathryn. 1993. "Factors Influencing Probation Outcome: A Review of the Literature," *Federal Probation* 57 (2): 23-29.
- Petersilia, Joan. 1987. "Probation and Felony Offenders," *Federal Probation* 51(2), 56-61.
- _____. 1997. " Probation in the United States," In Michael Tonry (Eds.). *Crime and Justice: An Annual Review of Research* 22: 149-200. Chicago, IL: University of Chicago Press.
- _____. 1999. "A Decade of Experimenting with Intermediate Sanctions: What Have We Learned?" *Justice Research and Policy* 1(1):9-23.

- _____. 2011. "Community Corrections: Probation, Parole, and Prisoner Reentry," In James Q. Wilson and Joan Petersilia (Eds.), *Crime and Public Policy*: 499-531. New York, NY: University of Oxford Press.
- Petersilia, Joan, and Susan Turner. 1993. "Intensive Probation and Parole," In Michael Tonry (Eds.). *Crime and Justice: An Annual Review of Research* 17: 281-335. Chicago, IL: University of Chicago Press.
- Ritter, Christian, and Martin A. Tanner. 1992. "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler," *Journal of the American Statistical Association* 87(419): 861-868.
- Robinson, David. 1995. *The Impact of Cognitive Skills Training on Post-Release Recidivism among Canadian Federal Offenders*. Ottawa, Ontario: Correctional Research and Development, Correctional Service Canada.
- Rosenfeld, Richard, Joel Wallman, and Robert Fornango. 2005. "The Contribution of Ex-Prisoners to Crime Rates," In Jeremy Travis and Christy Visser (Eds.). *Prisoner Reentry and Crime in America*. New York: Cambridge University Press.
- Ross, Robert R., Elizabeth A. Fabiano, and Crystal D. Ewles. 1988. "Reasoning and Rehabilitation," *International Journal of Therapy and Comparative Criminology* 32(1): 29-35.
- Rossi, Peter. 1987. "The Iron Law Of Evaluation And Other Metallic Rules," *Research in Social Problems and Public Policy* 4: 3-20.
- Roundtree, George A., Dan W. Edwards, and Jack B. Parker. 1984. "A Study of Personal Characteristics Related to Recidivism," *Journal of Offender Counseling Services and Rehabilitation* 8(3): 53-61.
- Scott, Menard, and Herbert Covey. 1983. "Community Alternatives and Rearrest in Colorado," *Criminal Justice and Behavior* 10(1): 93-108.
- Shpitser, Ilya, and Judea Pearl. 2006. "Identification of Conditional Interventional Distributions," In Rina Dechter, and Thomas S. Richardson (Eds.), *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial intelligence*: 437-444. Corvallis, OR: AUAI Press.
- Smith, Paula, Paul Gendreau, and Kristin Swartz. 2009. "Validating the Principles of Effective Intervention: A Systematic Review of the Contributions of Meta-Analysis in the Field of Corrections," *Victims and Offenders* 2(4): 148-169.
- Sommer, Alfred, and Scott Zeger. 1991. "On Estimating Efficacy from Clinical Trials," *Statistics in Medicine* 10: 45-52.
- Spruance, Lisa M., Patricia Van Voorhis, Shelley J. Listwan, Jennifer Pealer, and Renita Seabrook, n.d. *The Georgia Cognitive Skills Experiment: Process Evaluation Phrase Two*. Extracted from

- [http://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/Georgia Phase 2 process eval.pdf](http://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/Georgia_Phase_2_process_eval.pdf) being accessed on Jan. 12, 2014.
- Sullivan, Christopher J., and Thomas Mieczkowski. 2008. "Bayesian Analysis and the Accumulation of Evidence in Crime and Justice Intervention Studies," *Journal of Experimental Criminology* 4(4): 381-402.
- Van Voorhis, Patricia, Lisa M. Spruance, P. Neal Ritchey, Shelley J. Listwan, and Renita Seabrook. 2004. "The Georgia Cognitive Skills Experiment: A Replication of Reasoning and Rehabilitation," *Criminal Justice and Behavior* 31:282-305.
- Weisburd, David, Cynthia M. Lum, and Sue-Ming Yang. 2003. "When Can We Conclude That Treatments or Programs 'Don't Work'?" *Annals of the American Academy of Political and Social Science* 587: 31-48.
- Wilson, David B., Leana A. Bouffard, and Doris L. MacKenzie. 2005. "Quantitative Review of Structure, Group-Oriented, Cognitive-Behavioral Programs for Offenders," *Criminal Justice and Behavior* 32:172-204.
- Wood, Yvonne R., and Clifford R. O'Donnell. 1980. "Adult Probation Adjustment: A Twelve Month Follow-up," *Journal of Community Psychology* 9(1): 86-92.